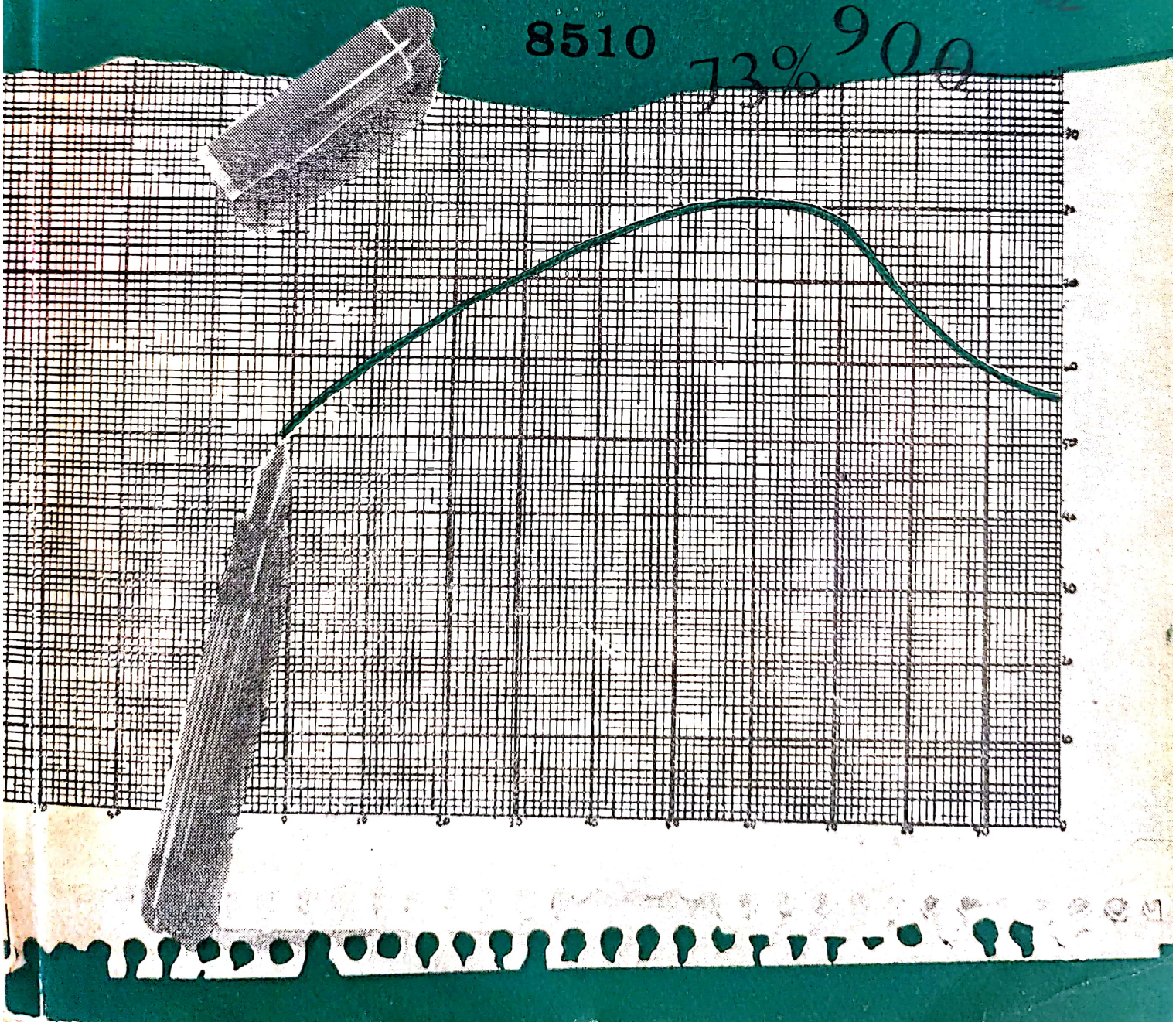
V. GIURGIU
8387

APLICAȚII ALE STATISTICII MATEMATICE în SILVICULTURĂ

62
8510
73%900



Scanned with OKEN Scanner

Dr. ing. V. GIURGIU

APLICAȚII ALE STATISTICII MATEMATICE în SILVICULTURĂ

Scanned with OKEN Scanner

1

INTRODUCERE

1.1. GENERALITĂȚI

Statistica matematică cuprinde un complex de metode care permit să se obțină din datele de observație și experiment concluzii fundamentate din punct de vedere științific, asupra fenomenelor de masă în care intervine variabilitatea statistică.

Aceste metode se aplică în toate domeniile unde se întîlnesc asemenea fenomene: biologie, agricultură, chimie, fizică, industrie etc. In economia forestieră, statistica matematică are un cîmp larg de aplicare, fapt explicabil dacă avem în vedere că majoritatea fenomenelor luate în considerare au un pronunțat caracter de masă. In silvicultură nu putem controla în suficientă măsură toți factorii care intervin în desfășurarea experimentelor, mai ales că unele rezultate se obțin după expirarea unor perioade foarte lungi, uneori de ordinul zecilor de ani. In aceste condiții intervine variabilitatea individuală: de la sămînță la sămînță, de la arbore la arbore, de la arboret la arboret, de la un tip la alt tip de pădure, de la moment la moment etc. Desigur, aceste fluctuații, care de cele mai multe ori au un caracter întîmplător, îngreuiază interpretarea rezultatelor experimentării sau ale observațiilor, existînd întotdeauna o oarecare probabilitate (șansă) ca un rezultat (efect) constatat să fie datorit întîmplării. Statistica matematică, axîndu-se pe teoreme ale calculului probabilităților, permite măsurarea acestei șanse, avînd ca bază variabilitatea specifică fenomenului studiat. Fără o prelucrare adecvată a rezultatelor nu putem ști în ce măsură o diferență (dintre două sau mai multe experiențe) este semnificativă; nu ne putem da seama de gradul de legătură dintre două (sau mai multe) fenomene ori caracteristici. Statistica matematică ajută, în multe situații, ca dintr-o serie de observații cu un pronunțat caracter contradictoriu să se tragă o concluzie; înlătură riscul unor concluzii insuficient fundamentate. Totodată devine un instrument indispensabil și în organizarea experimentărilor, arătînd, spre exemplu, care este numărul minim de măsurători pentru a garanta o experiență concludentă. In acest mod aplicarea metodelor statistico-matematice contribuie la obținerea unui maximum de informație științifică cu minimum de cheltuială.

Scanned with OKEN Scanner

O colectivitate (*populație*) reprezintă *o totalitate de unități calitativ omogene, cu un număr de însușiri esențiale comune*: totalitatea arborilor dintr-un arboret, totalitatea suprafețelor de probă amplasate într-un masiv păduros omogen, totalitatea semințelor recoltate dintr-un arbore etc. Dar, o colectivitate poate fi nu numai o populație de indivizi (spre exemplu — arboretul), ci și o totalitate de numărători, măsurători și valori experimentale; este însă necesar ca aceste elemente să aibă anumite trăsături comune.

În cercetările din domeniul silviculturii de cele mai multe ori se folosește un material relativ restrîns, obținut prin observație sau pe cale experimentală. Acest material formează o *colectivitate de selecție*, care reprezintă deci o parte dintr-o populație, extrasă în scopul caracterizării fenomenului în ansamblul său. Prin intermediul acestei colectivități de selecție (eșantion) se tinde spre cunoașterea colectivității generale. Arborii de probă aleși și studiați în anumite scopuri reprezintă un eșantion, ca și loturile experimentale instalate în vederea comportării anumitor clone de plopi euramericani. Un eșantion reprezintă și totalitatea epruvetelor recoltate în vederea studierii proprietăților fizico-mecanice ale lemnului anumitor specii forestiere.

Este deosebit de important să ne dăm seama că în silvicultură aproape toate observațiile și experimentările ce se fac se bazează pe colectivități de selecție. În cadrul majorității acestor colectivități de selecție rezultatele nu sînt identice de la observație la observație, de la o *unitate statistică* la altă unitate, fapt explicabil dacă avem în vedere intervenția unor factori întîmplători, precum și din cauza variabilității intrinsece a fenomenelor, mai ales a celor biologice. De aici rezultă larga aplicabilitate a metodelor statisticii matematice în silvicultură.

În silvicultură tratarea majorității problemelor, cu legile degajate și previziunile corespunzătoare dobîndește un caracter *probabilist*. De aceea, orice generalizare trebuie să se bazeze pe considerente de ordin *statistic*, pentru a avea un caracter științific. Faptul acesta este adevărat de la cele mai elementare determinări biometrice pînă la generalizări referitoare la ecologia speciilor forestiere sau la manifestările complexe ale biogeocenozei forestiere.

Se știe din teoria probabilităților, pe care se sprijină calculele statistice, că numai un eșantion destul de numeros și suficient de reprezentativ poate scoate în evidență tendințele principale, aspectele cele mai frecvente, arată posibilitatea ce există ca o manifestare anumită să fie întîlnită în colectivitatea considerată. În acest mod, afirmația se poate prezenta alături de un coeficient de probabilitate, ceea ce are o importanță deosebită atît din punct de vedere științific, cît și în aplicațiile practice. În plus, tratarea statistică permite să se aprecieze și eroarea probabilă a afirmațiilor, deci limitele între care trebuie să se admită un joc imprevizibil al rezultatelor (limitele fiind însă prevăzute).

În legătură cu valabilitatea aplicării metodelor statisticii matematice în silvicultură trebuie însă să atragem atenția asupra faptului că, în cercetarea fenomenelor de masă întîlnite în acest domeniu, apar probleme concrete care necesită metode specifice de cercetare, precum și o analiză temeinică a *legăturilor cauzale* respective. Se impune principiul analizei calitative a

Scanned with OKEN Scanner

fenomenelor înaintea celei cantitative. Procedînd astfel, evităm tendința devierilor *formaliste* în aplicarea acestor metode. Totodată, trebuie evitată tendința devierilor „*probabiliste* care — în fața complexității unui fenomen — renunță la analiza și clasificarea importanței diferiților factori, preferînd cercetarea formală a fenomenului în termenii statisticii matematice" (S ă h l e a n u, 1957).

Aplicînd aceste metode în silvicultură, cercetătorul nu va întîrzia deci să facă o prealabilă analiză calitativă a fenomenelor și a proceselor studiate și nu va pierde din vedere nici faptul că metodele matematice reprezintă un instrument auxiliar al cercetării de *netăgăduită utilitate.* Mai mult decît atît, aplicarea metodelor statisticii matematice în silvicultură este nu numai utilă dar și indispensabilă. Fără aceste metode, variabilitatea specifică majorității fenomenelor și caracteristicilor studiate de silvicultor, nu poate fi „măsurată" sau luată în considerare la testarea anumitor diferențe, la formularea anumitor concluzii. Multe fenomene, deși sînt studiate de multă vreme, au rămas în mare măsură neclarificate tocmai din cauza folosirii insuficiente a metodelor statisticii matematice. În această situație se găsește, spre exemplu, problema influenței intensității răriturilor asupra productivității totale a arboretului.

Pentru silvicultor o importanță deosebită o au procedeele statisticii matematice oriunde intervine observația și experimentul, înlesnindu-i trecerea de la observație la concluzii științific fundamentate. Cunoașterea acestor procedee este necesară nu numai silvicultorului cercetător, dar și silvicultorului practician, care urmărește o analiză critică a literaturii de specialitate în vederea alegerii concluziilor și recomandărilor riguros fundamentate.

Desigur, metodele statisticii matematice sînt inegal de importante în diferitele domenii de activitate ale silvicultorului. Deosebit de utile sînt pentru biometria forestieră, mai ales în dendrometrie și auxometrie, în auxonomie, precum și în cercetările privind cunoașterea caracteristicilor fizice, mecanice și chimice ale lemnului. Dar nici cercetările din domeniul silvobiologiei nu mai pot fi finalizate fără o temeinică fundamentare a concluziilor prin procedeele statisticii matematice, îndeosebi prin aplicarea analizei dispersionale. Avem în vedere în primul rînd problemele eredității, precum și cele privind ecologia speciilor forestiere. De o utilitate incontestabilă este aplicarea teoriei corelației și analizei dispersionale în toate cercetările privind silvotehnica (tehnica împăduririlor, culturi comparative, îngrijirea arboretelor, regenerarea pădurilor, ameliorarea stațiunilor etc.). Studiile tipologice, de asemenea, se cer a fi susținute prin procedee ale statisticii matematice, fapt ce va ușura trecerea de la observație la anumite caracterizări tipologice științific fundamentate. Prin aceasta, statistica matematică poate imprima mai multă rigurozitate tipologiei forestiere. Fără aplicarea procedeelor statisticii matematice este dificil de organizat controlul calității producției forestiere. Nici problemele de economie forestieră, cum sînt cele privind asigurarea eficienței economice maxime prin reducerea prețului de cost, realizarea beneficiului maxim, cheltuieli minime de muncă socială etc. nu mai pot fi rezolvate fără promovarea perseverentă a metodelor

Scanned with OKEN Scanner

matematice, în special a metodelor matematice moderne (programarea matematică, teoria jocurilor matematice, teoria firelor de așteptare etc.).

Aplicarea pe scară largă a metodelor matematice moderne în economia forestieră, mai ales pentru soluționarea problemelor de mare serie sau pentru rezolvarea unor probleme de complexitate deosebită, în noile condiții este legată de folosirea unei tehnici de calcul de mare randament. Aceasta înseamnă folosirea practică a calculatoarelor electronice.

1.2. OBSERVAREA ȘI GRUPAREA STATISTICĂ

O cercetare statistică tipică trece prin trei etape distincte:

— observarea și recoltarea informațiilor necesare;

— stratificarea (gruparea), centralizarea și prelucrarea informațiilor statistice;

— analiza, interpretarea și prezentarea rezultatelor obținute.

Observarea statistică are o importanță hotărîtoare asupra reușitei cercetării statistice întreprinse; de aceea, în cadrul acestei etape trebuie asigurate condițiile cerute de orice cercetare științifică, și anume: competența, sinceritatea și imparțialitatea cercetătorului, precum și uniformitatea criteriilor de măsurare, notare sau de selecționare a materialului înregistrat.

Tot atît de important este momentul definirii precise a scopului și a sarcinilor cercetării statistice, indicarea clară a obiectului observării și a unităților de observare, precum și limitarea în timp și spațiu a lucrărilor.

După etapa observării statistice — desfășurată diferențiat în funcție de esența și specificul fenomenelor supuse cercetării — urmează centralizarea datelor. La baza centralizării stă *stratificarea* (gruparea) științifică a materialului — operație de o importanță deosebită. Stratificarea are rolul de a delimita grupe *calitativ omogene* în cadrul totalității unităților care constituie obiectul cercetării. Îndeplinirea acestei condiții conduce nemijlocit la formarea unor *colectivități* (*populații*) statistice, constituite din unități identice din punct de vedere calitativ. Dăm mai jos cîteva exemple de asemenea colectivități:

— totalitatea de arbori de aceeași specie, vîrstă, proveniență și crescuți în aceleași condiții staționale (în cadrul unui arboret format din mai multe specii, cu arbori de vîrste și proveniențe diferite, în scopul unui studiu statistic științific fundamentat, se impune o *stratificare* a arborilor în raport cu specia, proveniența și vîrsta);

— totalitatea puieților de aceeași specie, vîrstă și proveniență, plantați în condiții staționale identice;

— bușteni pentru gater proveniți din același arboret;

— cheresteaua obținută la același gater într-o anumită perioadă de timp, din material avînd aceeași proveniență și calitate.

Numai prin astfel de grupări devine posibilă caracterizarea științifică a întregii colectivități studiate, în raport cu tipurile ei calitative.

Scanned with OKEN Scanner

Folosind caracteristici considerate tipice sau esențiale din anumite puncte de vedere, prin grupări statistice adecvate, se poate ajunge la clasificări *tipologice*. Asemenea clasificări pot avea valoare științifică „numai atunci cînd sînt formulate anumite scopuri de cunoaștere, motivate de teorie; cînd se iau, drept caracteristici ale grupărilor, caracteristici esențiale determinate de teorie" (S ă h l e a n u, 1957).

1.3. ÎNREGISTRAREA ȘI PRELUCRAREA INIȚIALĂ A DATELOR

Rezultatele observării se înregistrează în fișe cuprinzînd *datele inițiale*. Acestea se pot referi la una sau la mai multe însușiri sau caracteristici ale unităților observate. În tabelul 1.1 sînt prezentate rezultatele măsurării creșterii în diametru pe 5 ani la 192 arbori dintr-un arboret de molid de 50 ani.

Tabelul 1.1

Creșterea în diametru pe 5 ani la 192 arbori într-un arboret echien de molid de 50 ani în mm (Giurgiu, 1957)

9	6	4	7	6	6	9	5	10	7	9	9
10	4	7	11	6	1	5	6	8	15	4	2
5	6	4	8	5	2	5	8	7	7	6	8
.											.
.											.
.											.
.											.
5	4	7	12	8	9	10	13	7	7	10	8

Pentru a scoate în evidență caracterul distribuției arborilor în raport cu creșterea lor în diametru și pentru a ușura prelucrarea statistică ulterioară, se formează *clase*, grupînd valorile ce aparțin unui interval dat.

Numărul intervalelor poate varia, în funcție de condițiile concrete ale fenomenului studiat; de obicei, numărul claselor nu trebuie să depășească 15—20, dar se impune ca acesta să fie mai mare de 5. Frecvent, amplitudinea de variație se împarte în aproximativ 10 intervale. Un număr prea mare de intervale îngreuiază prelucrarea statistică ulterioară; totodată nu poate fi scos în evidență cu suficientă claritate caracterul distribuției. Un număr prea mic de intervale prezintă dezavantaje din punct de vedere al preciziei.

Pentru determinarea mărimii intervalului (h) se alege cea mai mare (x_{max}) și cea mai mică (x_{min}) valoare a variabilei x și diferența dintre ele se împarte la numărul n de intervale ales:

$$h = \frac{x_{max} - x_{min}}{n} \qquad (1.1)$$

Scanned with OKEN Scanner

Pentru exemplul din tabelul 1.1 obținem

$$h = \frac{15 - 1}{10} = 1{,}4 \cong 1 \text{ mm}$$

Datele inițiale se prezintă sub formă tabelară, după operația de *despuiere*. Aceasta constă în repartizarea unităților statistice în raport cu mărimea variabilei x, fiecare unitate observată fiind înlocuită prin semne convenționale: linii (tabelul 1.2) sau puncte. Pentru fiecare clasă se stabilesc, după

Tabelul 1.2

Despuierea datelor din tabelul 1.1. și calculul frecvențelor absolute, frecvențelor relative, frecvențelor absolute cumulate și al frecvențelor relative cumulate

Mărimea variabilei x_i (creșterea în diametru, mm)	Punctaj	Numărul observațiilor (frecvența absolută) n_i	Frecvența absolută cumulată	Frecvența relativă	Frecvența relativă cumulată
1*	II	2	2	0,011	0,011
2	IIII IIII II	12	14	0,063	0,074
3	IIII IIII IIII II	17	31	0,089	0,163
4	IIII IIII IIII IIII	19	50	0,099	0,262
5	IIII IIII IIII IIII IIII I	31	81	0,161	0,423
6	IIII IIII IIII IIII II	22	103	0,114	0,537
7	IIII IIII IIII IIII IIII	24	127	0,125	0,662
8	IIII IIII IIII IIII III	23	150	0,119	0,781
9	IIII IIII III	13	163	0,068	0,849
10	IIII IIII II	12	175	0,062	0,911
11	IIII III	8	183	0,042	0,953
12	IIII	5	188	0,026	0,979
13	II	2	190	0,011	0,990
14		0	190	0	0,990
15	II	2	192	0,010	1,000
Total	—	192	—	1,000	—

* Cuprinde valori de la 0,6 la 1,5, întocmai cum în următoarea clasă sînt cuprinse valori de la 1,6 la 2,5 ș.a.

cum se poate urmări în tabelul 1.2, următorii indicatori: frecvența absolută (n), adică numărul observațiilor corespunzătoare fiecărei clase; frecvența absolută cumulată; frecvența relativă, adică raportul dintre frecvența absolută și numărul total al observațiilor; frecvența relativă cumulată. Pe plan probabilistic, frecvența relativă este o estimație a probabilității *, iar noțiunii de frecvență cumulată îi corespunde noțiunea de *funcție de distribuție*. Prin raportarea frecvențelor relative la mărimea intervalului (h) se obține *densitatea distribuției*. În loc de frecvențe relative cumulate se folosește noțiunea de *funcție de distribuție*, iar pentru seria frecvențelor relative se utilizează denumirea de *funcție de frecvență*.

* Probabilitatea unui eveniment care se produce la întîmplare este raportul dintre numărul cazurilor favorabile și numărul cazurilor posibile cu condiția ca toate cazurile să fie egal posibile.

Scanned with OKEN Scanner

În situaţia unor lucrări statistice de amploare, cu un număr mare de unităţi observate, fiecare unitate fiind studiată din mai multe puncte de vedere, despuierea „manuală" necesită mult timp. De aceea se foloseşte fie „metoda semimanuală", bazată pe fişe perforate marginal, fie „metoda mecano-grafică" a cartelelor perforate şi a maşinilor tip HOLLERITH (P r o d a n, 1961).

În ultimul timp se folosesc tot mai mult calculatoarele electronice.

1.4. REPREZENTĂRI GRAFICE ALE DISTRIBUŢIILOR STATISTICE

Seriile de variaţie (distribuţie), cum sînt cele obţinute în tabelul 1.2, în scopul de a pune mai clar în evidenţă caracterul variaţiei fenomenului cercetat, se reprezintă grafic sub formă de *histograme* (fig. 1.1 şi 1.2) şi *poligoane* (fig. 1.3). În cazul a două variabile se construiesc, de obicei, *prizmograme* (fig. 1.4).

Reprezentarea grafică prin intermediul poligoanelor sau a curbelor de frecvenţă este posibilă numai în cazul *distribuţiilor de tip continuu*, cînd variabila poate lua orice valoare într-un interval dat. Distribuţiile de *tip discret* (variabila x poate lua în orice interval finit un număr finit de valori) pot fi reprezentate grafic prin intermediul histogramelor.

Distribuţiile de o caracteristică calitativă (de pildă, repartiţia fondului forestier pe specii) se reprezintă grafic prin *diagrame*. Deosebit de ilustrative sînt *diagramele de structură*, în care frecvenţele sînt reprezentate prin dreptunghiuri sau sectoare de cerc, unde înălţimea dreptunghiurilor sau numărul de grade al arcelor corespunzătoare sînt proporţionale cu mărimea frecvenţelor (fig. 1.5).

Distribuţiile teritoriale se reprezintă prin intermediul *cartogramelor* (vezi planşa) sau al *cartodiagramelor* (fig. 1.6), pe care datele statistice sînt reprezentate prin haşuri, puncte sau culori de intensitate mai mare sau mai mică (în cazul cartogramelor), ori prin diferite figuri geometrice (în cazul cartodiagramelor).

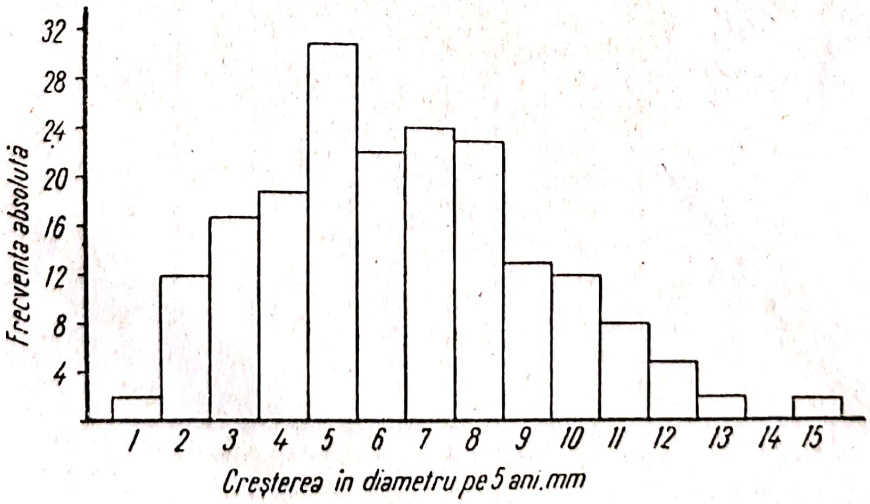
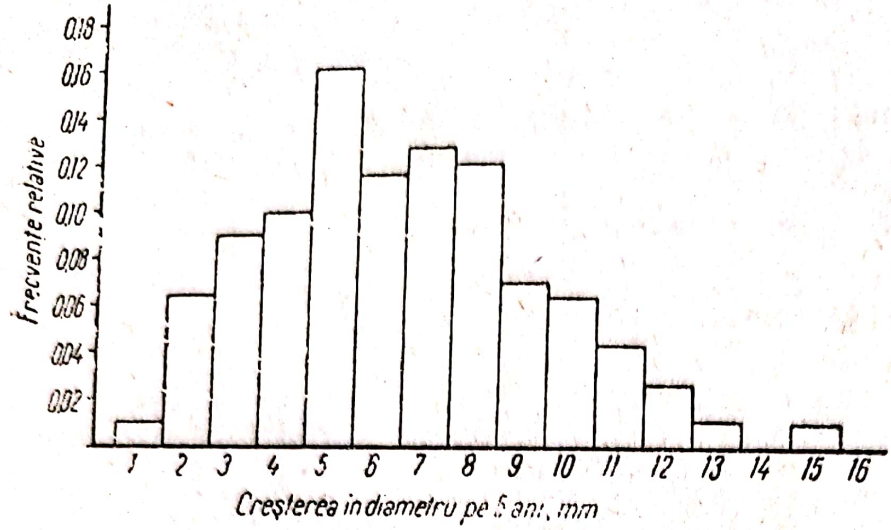
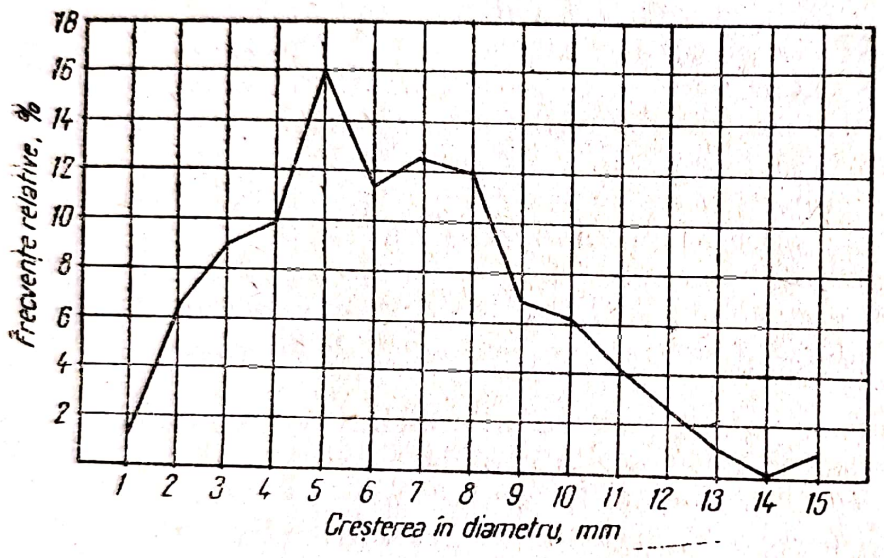


Fig. 1.1. Histograma frecvenţelor absolute, construită pe baza datelor din tabelul 1.2

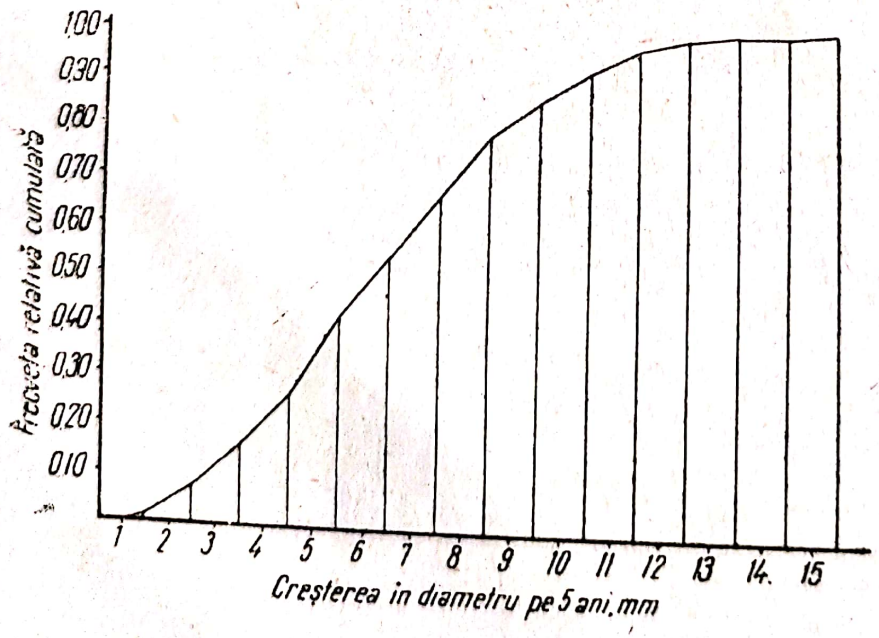
Scanned with OKEN Scanner



← Fig. 1.2. Histograma frecvențelor relative, construită pe baza datelor din tabelul 1.2



← Fig. 1.3. *a*. Poligonul frecvențelor relative, construit pe baza datelor din tabelul 1.2



← Fig. 1.3. *b*. Poligonul frecvențelor relative cumulate, construit pe baza datelor din tabelul 1.2

Scanned with OKEN Scanner

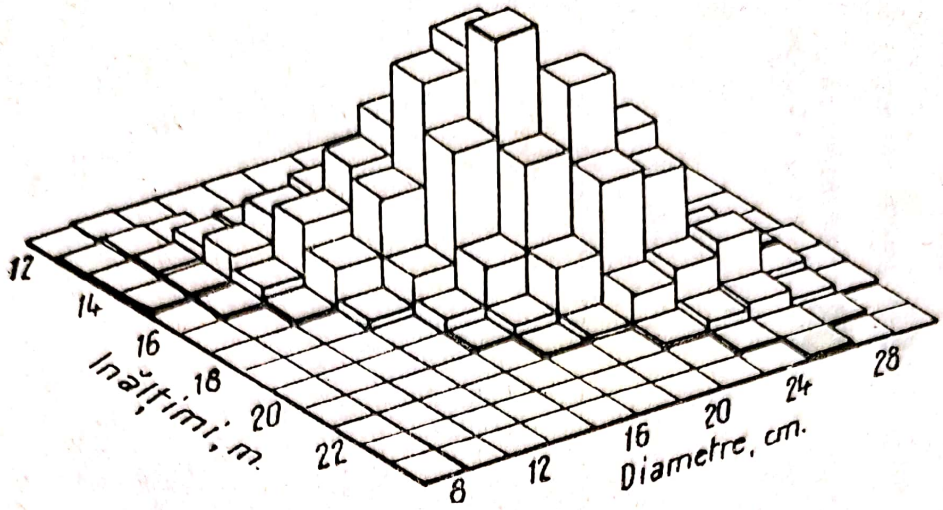


Fig. 1.4. Prizmogramă (repartiția arborilor în raport cu diametrele și înălțimile lor într-un arboret de molid)

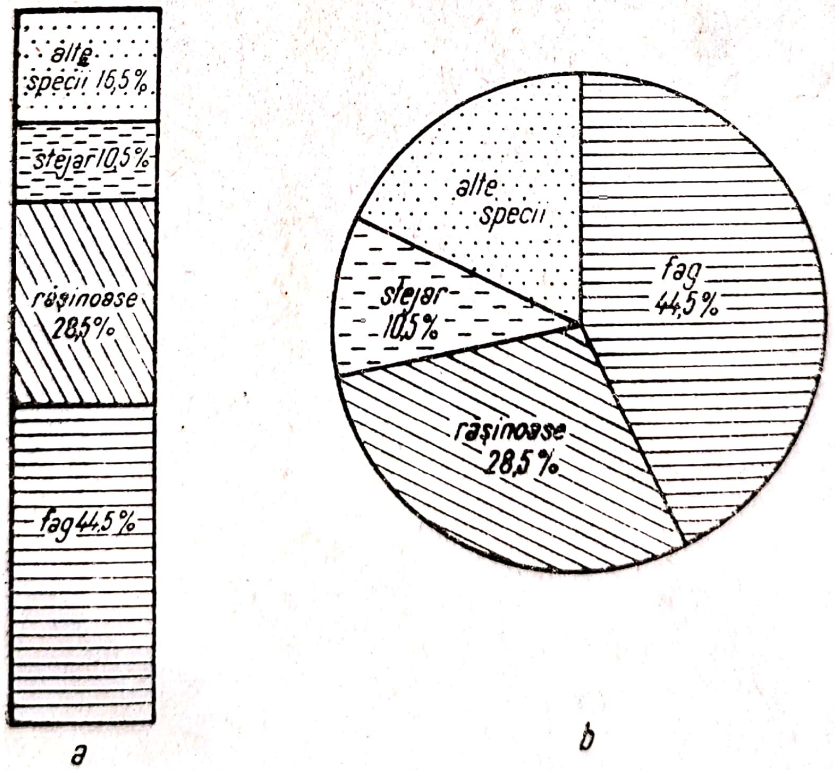


Fig. 1.5. Diagrama de structură: *a*) în dreptunghi; b) în cerc (ponderea pe specii a masei lemnoase exploatată în anul 1964)

Scanned with OKEN Scanner

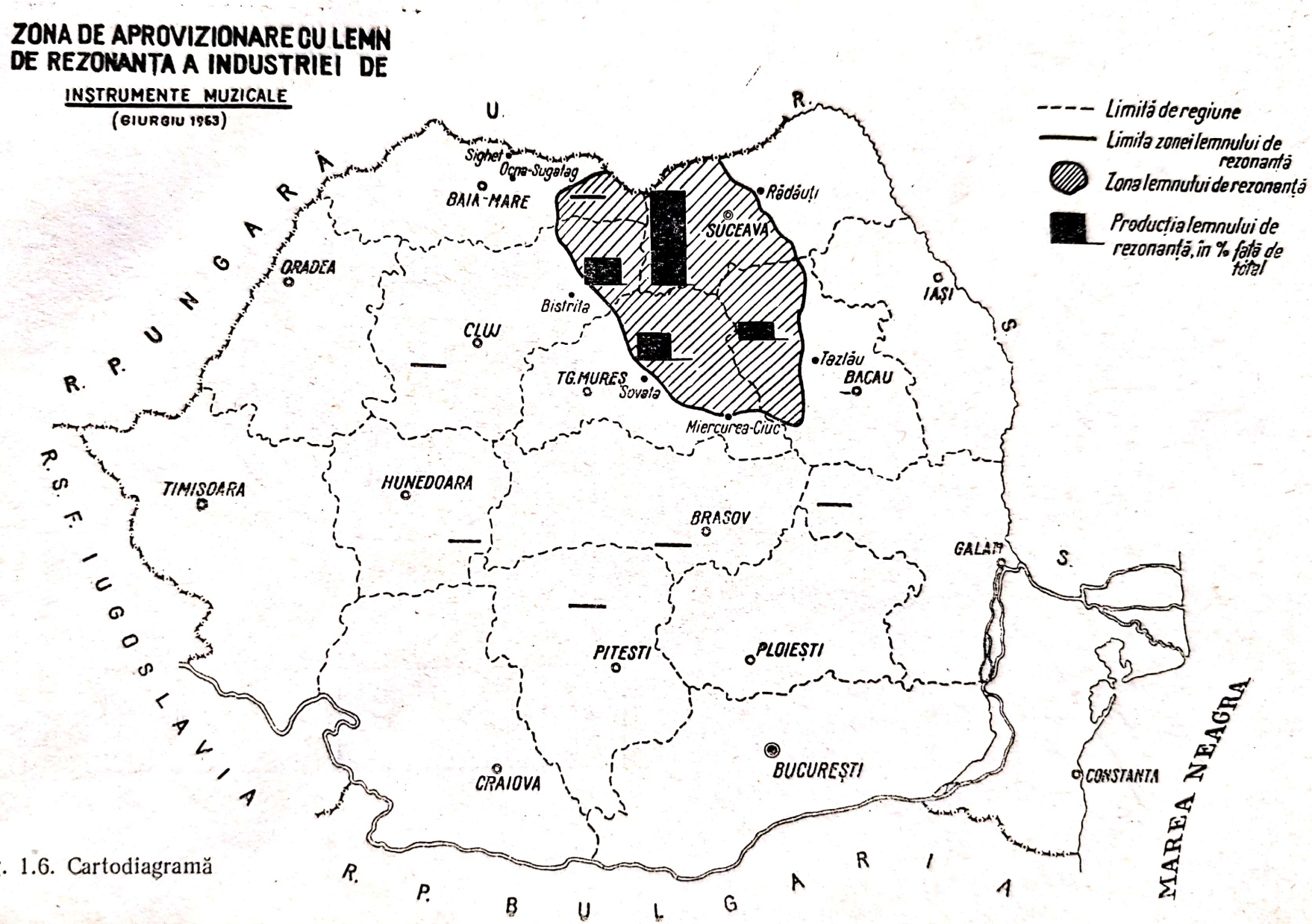


Fig. 1.6. Cartodiagramă

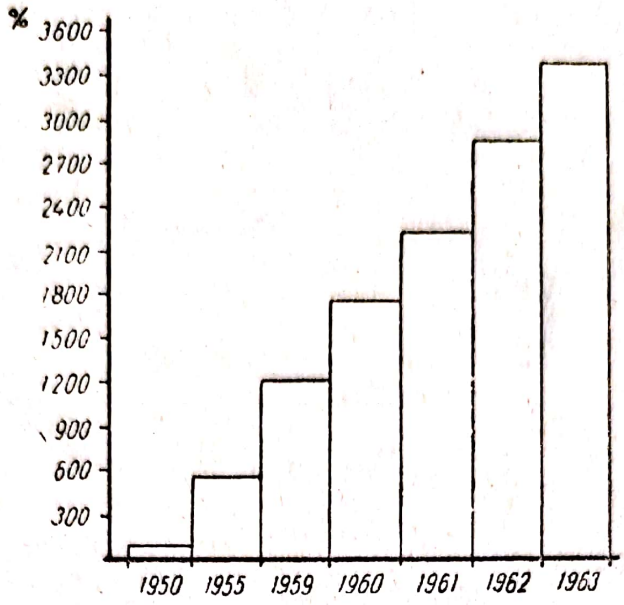


Fig. 1.7. Historiogramă (dinamica producției de mobilă în R.S. România) (anul de bază 1950)

Seriile care caracterizează un anumit fenomen în diferite momente succesive ale timpului — așa-numitele serii de momente — pot fi redate grafic prin *historiograme* (fig. 1.7).

Scanned with OKEN Scanner

II

MĂRIMI MEDII

2.1. GENERALITĂȚI

Un prim pas spre prezentarea sintetică a datelor observării statistice îl constituie evidențierea *distribuțiilor statistice* (seriilor de variație), așa cum s-a arătat în capitolul anterior (tabelul 1.2). Reducerea în continuare a volumului de date se realizează prin calcularea *mărimilor medii*. Unele dintre aceste mărimi se referă la *poziția* distribuției (mediile, mediana, modulul), altele definesc *variabilitatea* unităților statistice (întinderea cîmpului de variație, abaterea medie pătratică, coeficientul de variație etc.). Utili sînt și parametrii care reflectă *forma* distribuției (indicele asimetriei, indicele excesului etc.).

Înainte de a trece la calculul mărimilor medii se impune o atentă și multilaterală analiză a materialului recoltat, spre a obține garanția *omogenității* lui. Folosind valorile medii la analiza fenomenelor studiate, permanent trebuie avut în vedere că valoarea medie este întotdeauna *media* multor valori diferite de același tip și că mediile generale și neîntemeiate au o însemnătate cu totul fictivă.

Din practica și cercetarea forestieră se pot da multe exemple de folosire a unor medii generale și neîntemeiate capabile să reflecte denaturat unele fenomene. Subliniem astfel inutilitatea: determinării unor vîrste medii la arboretele pluriene, formate din mai multe colectivități statistice; stabilirii unei înălțimi medii globale pentru arboretele bietajate; fixării unui ciclu de producție mediu pentru unități de producție neomogene din punct de vedere la compoziției, al condițiilor naturalistice și al obiectivelor economice urmărite; stabilirii unor norme medii pentru lucrări neomogene sub aportul complexității lor etc.

2.2. MOMENTE

Mărimile medii folosite în practica statisticii reprezintă cazuri particulare ale unei noțiuni generale: *momente ale distribuțiilor statistice*. Acestea permit o determinare comodă a mediei, a abaterii medii pătratice, a indicelui

Scanned with OKEN Scanner

de asimetrie și a altor mărimi medii statistice. De aceea, în cele ce urmează prezentăm formulele pentru calculul lor.

Momentele reprezintă medii aritmetice ale abaterilor — ridicate la diferite puteri — de la un anumit punct al distribuției statistice, ales ca origine.

În funcție de punctul ales ca origine, se deosebesc mai multe categorii de momente.

Momentele obișnuite se calculează în raport cu o origine aleasă arbitrar. Expresia matematică a momentelor obișnuite este următoarea:

$$m_K = \frac{\Sigma (x_i - x_0)^K n_i}{\Sigma n_i} \tag{2.1}$$

unde:

m_K reprezintă momentul obișnuit de ordinul K, unde $K = 0, 1, 2, 3, \ldots, n$
x_i — valori observate;
x_o — varianta aleasă ca origine;
n_i — frecvențele corespunzătoare variantelor x_i;
K — ordinul momentului.

Dînd diferite valori pentru K obținem momente de diferite ordine.

Momentele centrale reprezintă medii aritmetice ale abaterilor — ridicate la diferite puteri — de la media aritmetică aleasă ca origine. Calculul acestora se face după următoarea formulă generală:

$$\mu_K = \frac{\Sigma (x_i - \bar{x})^K n_i}{\Sigma n_i} \tag{2.2}$$

Momentele centrale definesc majoritatea indicatorilor folosiți în operațiile statistico-matematice. Calculul lor este însă complicat și necesită stabilirea prealabilă a mediei aritmetice. Cu mult mai ușor se calculează momentele obișnuite, deoarece abaterile se stabilesc față de o bază arbitrară; de obicei se operează cu cifre întregi. De aceea prezintă interes practic următoarele formule de trecere de la momentele obișnuite la momentele centrale:

$$\mu_o = m_o = 1 \tag{2.3}$$

$$\mu_1 = o \tag{2.4}$$

$$\mu_2 = m_2 - m_1^2 \tag{2.5}$$

$$\mu_3 = m_3 - 3\ m_2 m_1 + 2\ m_1^3 \tag{2.6}$$

$$\mu_4 = m_4 - 4\ m_3 m_1 + 6\ m_2 m_1^2 - 3\ m_1^4 \tag{2.7}$$

Calculul momentelor obișnuite se poate efectua prin două procedee: procedeul direct și procedeul sumelor.

Calculul momentelor prin procedeul direct este exemplificat în tabelul 2.1.

Ultimele două coloane servesc pentru verificarea corectitudinii calculelor efectuate. Coloana a opta se obține adăugînd o unitate la abaterile ξ_i. Cifrele din ultima coloană reprezintă produsul dintre frecvențele n_i și abaterile ξ_i ridicate la puterea a patra. Pentru varianta x_1 obținem:

$$\xi_i + 1 = -5 + 1 = -4; \qquad (\xi_1 + 1)^4 n_1 = -4^4 \cdot 2 = 512$$

Tabelul 2.1

Calculul momentelor obișnuite prin *procedeul direct*, pentru o serie statistică reprezentînd variația modulului elasticității la compresiune paralel cu fibrele la molidul de rezonanță

x_i	n_i	$\xi_i = \frac{x_i - x_o}{h}$	$\xi_i n_i$	$\xi_i^2 n_i$	$\xi_i^3 n_i$	$\xi_i^4 n_i$	$\xi_i + 1$	$(\xi_i + 1)^4 n_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
60 000	2	—5	—10	50	—250	1 250	—4	512
74 000	5	—4	—20	80	—320	1 280	—3	405
88 000	20	—3	—60	180	—540	1 620	—2	320
102 000	44	—2	—88	176	—352	704	—1	44
116 000	87	—1	—87	87	— 87	87	0	0
x_o = 130 000	100	0	0	0	0	0	1	100
144 000	79	1	79	79	79	79	2	1 264
158 000	42	2	84	168	—336	672	3	3 402
172 000	16	3	48	144	432	1 296	4	4 096
186 000	4	4	16	64	256	1 024	5	2 500
200 000	1	5	5	25	125	625	6	1 296
	400	—	—33	1 053	—321	8 637	—	13 939

În scopul simplificării calculelor se prezintă în anexa 1 valorile expresiei

$$(\xi_i + 1)^4 n_i$$

pentru n_i de la 1 la 50 și pentru $(\xi_i + 1)$ de la 2 la 7.

Folosind formula (2.1) obținem momentele obișnuite:

$$m_1 = \frac{\Sigma \xi_i n_i}{\Sigma n_i} = \frac{-33}{400} = -0{,}0825; \qquad m_2 = \frac{\Sigma \xi i^2 n_i}{\Sigma n_i} = \frac{1\,053}{400} = 2{,}6325$$

$$m_3 = \frac{\Sigma \xi_i^3 n_i}{\Sigma n_i} = \frac{-321}{400} = -0{,}8025; \qquad m_4 = \frac{\Sigma \xi_i^4 n_i}{\Sigma n_i} = \frac{8\,637}{400} = 21{,}5925$$

Pentru verificarea calculelor aplicăm ecuația lui CHARLIER:

$$m_4^* = m_0 + 4m_1 + 6m_2 + 4m_3 + m_4 \tag{2.8}$$

unde:

$$m_4^* = \frac{\Sigma (\xi_i + 1)^4 n_i}{\Sigma n_i} \tag{2.9}$$

Numărătorul acestei expresii reprezintă suma cifrelor trecute în coloana 9 a tabelului 2.1.

Prin urmare:

$$m_4^* = \frac{13\,939}{400} = 34{,}8475$$

După formula (2.8) obținem același rezultat.

Scanned with OKEN Scanner

Așa dar, calculul momentelor obișnuite este corect și — ca atare — putem trece la calculul momentelor centrale, după formulele (2.5) — (2.7):

$$\mu_2 = 2{,}6325 - (-0{,}0825)^2 = 2{,}6257$$

$$\mu_3 = -0{,}8025 - 3.2{,}6325\ (-0{,}0825) + 2\ (-0{,}0825)^3 = -0{,}152$$

$$\mu_4 = 21{,}5925 - 4\ (-0{,}8025)\ (-0{,}0825) + 6.2{,}6325)\ (-0{,}0825)^2 - \\ - 3\ (-0{,}0825)^4 = 21{,}432$$

Pentru verificarea calculelor se recomandă formulele:

$$\mu_3 = m_3 - 3\ \mu_2 m_1 - m_1^3 \qquad (2.10)$$

$$\mu_4 = m_4 - 4\ \mu_3 m_1 - 6\ \mu_2 m_1^2 - m_1^4 \qquad (2.11)$$

Folosind aceste formule, avem:

$$\mu_3 = -0{,}8025 - 3.2{,}6257.(-0{,}0825) - (-0{,}0825)^3 = -0{,}152$$

$$\mu_4 = 21{,}5925 - 4.(-0{,}152)(-0{,}0825) - 6.2{,}6257(-0{,}0825)^2 - \\ - (-0{,}0825)^4 = 21{,}432$$

Rezultatele fiind identice, stabilim că momentele centrale sînt calculate corect.

Pentru cazurile cînd variantele sînt grupate în clase, calculul momentelor este afectat de o oarecare eroare ca urmare a denaturării gradului de variabilitate a unităților statistice observate. De aceea trebuie aduse unele corecții, cunoscute sub denumirea de corecțiile SHEPPARD *. Aceste corecții se referă numai la momentele centrale $\mu_2, \mu_4, \mu_6 \ldots$ Momentele centrale corectate se stabilesc după formulele:

$$\mu_2' = \mu_2 - 0{,}083 \qquad (2.12)$$

$$\mu_4' = \mu_4 - 0{,}5\ \mu_2 + 0{,}029 \qquad (2.13)$$

În cazul nostru, obținem:

$$\mu_2' = 2{,}6257 - 0{,}083 = 2{,}5427$$

$$\mu_4' = 21{,}432 - 0{,}5.2{,}6257 + 0{,}029 = 20{,}409$$

În legătură cu acest mod de calcul facem mențiunea că rezultatele se obțin în unități convenționale, deoarece abaterile introduse în calcul au fost simplificate cu h — mărimea intervalului, fapt ce trebuie luat în considerare atunci cînd se trece la folosirea practică a momentelor.

Pentru momentele centrale exprimate în unități primare se procedează astfel:

$$\mu_2' = \mu_2 - 0{,}083\ h^2 \qquad (2.14)$$

$$\mu_4' = \mu_4 - 0{,}5\ \mu_2 h^2 + 0{,}029\ h^4 \qquad (2,15)$$

---

* Pentru seriile cu un număr relativ mare de intervale, corecțiile Sheppard, de obicei, nu se calculează.

208.872

Calculul momentelor prin procedeul sumelor este mult mai simplu, pentru care motiv poate fi frecvent folosit în practica calculului statistico-matematic. Potrivit acestui procedeu, înmulțirile sînt înlocuite prin simple adunări, calculele putîndu-se verifica pe parcurs. Folosirea lui se bazează pe faptul că între sumele S_i și diferențele d_i, obținute în cadrul acestui procedeu și momentele obișnuite m_i există relații perfect determinate, exprimate prin următoarele formule:

$$m_1 = \frac{d_1}{\Sigma n_i} \tag{2.16}$$

$$m_2 = \frac{S_1 + 2S_2}{\Sigma n_i} \tag{2.17}$$

$$m_3 = \frac{d_1 + 6\,d_2 + 6\,d_3}{\Sigma n_i} \tag{2.18}$$

$$m_4 = \frac{S_1 + 14\,S_2 + 36\,S_3 + 24\,S_4}{\Sigma n_i} \tag{2.19}$$

$$m_5 = \frac{d_1 + 30\,d_2 + 150\,d_3 + 240\,d_4 + 120\,d_5}{\Sigma n_i} \tag{2.20}$$

$$m_6 = \frac{S_1 + 62\,S_2 + 540\,S_3 + 1\,560\,S_4 + 1\,800\,S_5 + 720\,S_6}{\Sigma n_i} \tag{2.21}$$

Pentru determinarea sumelor S_i și a diferențelor d_i, folosite în formulele prezentate, se întocmește *tabela sumelor* (fig. 2.1).

x_i	n_i	(1)	(2)	(3)	(4)	(5)
15	2	2	2	2	2	2
14	0	2	4	6	8	10
13	2	4	8	14	22	32
12	5	9	17	31	53	85
11	8	17	34	65	118	
10	12	29	63	128		
9	13	42	105			
8	23	65				
7	24					
6	22	103				
5	31	81	178			
4	19	50	97	162		
3	17	31	47	65	85	
2	12	14	16	18	20	22
1	2	2	2	2	2	2

Fig. 2.1. Schema întocmirii tabelei sumelor

Spre exemplificare, vom calcula prin procedeul sumelor momentele distribuției prezentate în tabelul 1.2. Calculele sînt desfășurate în tabelul 2.2.

În prima coloană sînt înscrise în *ordine crescătoare* (*de jos în sus*) valorile variabilei x_i reprezentînd creșterea în diametru pe ultimii 5 ani, în milimetri. În coloana a doua sînt trecute frecvențele absolute corespunzătoare acestor variante. Intervalul dintre clase este $h = 1$ mm.

Prima operație constă în alegerea unei baze de pornire; aceasta se stabilește în mod arbitrar, fără a efectua calcule, de preferință la mijlocul amplitudinii de împrăștiere a valorilor observate. În exemplul luat, alegem $x_0 = 7$ mm.

Scanned with OKEN Scanner

Tabelul 2.2

Calculul momentelor obișnuite prin procedeul sumelor, pentru distribuția creșterilor în diametru la 192 arbori dintr-un arboret echien de molid

x_i	n_i	(1)	(2)	(3)	(4)	ξ_i+1	$(\xi_i+1)^4 n_i$
15	2	2	2	2	2	+9	13 122
14	0	2	4	6	8	+8	0
13	2	4	8	14	22	+7	4 802
12	5	9	17	31	53	+6	6 480
11	8	17	34	65	118	+5	5 000
10	12	29	63	128	—	+4	3 072
9	13	42	105	—	—	+3	1 053
8	23	65	—	—	—	+2	368
x_0=7	24	—	—	—	—	+1	24
6	22	103	—	—	—	+0	0
5	31	81	178	—	—	−1	31
4	19	50	97	162	—	−2	304
3	17	31	47	65	85	−3	1 377
2	12	14	16	18	20	−4	3 072
1	2	2	2	2	2	−5	1 250
	192	—	—	—	—	—	39 955
	a_i	170	233	246	203	$m_1=\frac{-111}{192}=-0{,}578$	
	b_i	281	340	247	107	$m_2=\frac{451+2\times573}{192}=8{,}319$	
	S_i	451	573	493	310	$m_3=\frac{-111+6(-107)+6(-1)}{192}=-3{,}953$	
	d_i	−111	−107	−1	96	$m_4=\frac{451+14\times573+36\times493+24\times310}{192}=175{,}318$	

A doua operație constă în trasarea semnelor de limitare a însumărilor, după următoarea regulă: în coloana (1) trasăm o linie în dreptul bazei x_0; în coloana (2) trasăm trei linii simetrice față de baza x_0; în coloana (3) trasăm 5 asemenea linii; în coloana (4) se trasează 7 linii ș.a.

După echiparea tabelei în acest mod, se procedează la înscrierea primei și ultimei frecvențe — neschimbate — în coloanele (1), (2), (3), (4) ș.a. În tabelul 2.2 frecvențele $n_1 = 2$ și $n_m = 2$ s-au înscris în coloanele respective. În continuare se trece la însumare, după schema prezentată în fig. 2.1. Astfel, cifrele coloanei (1), de sus în jos pînă la linia centrală, s-au obținut astfel: 2; 2 + 0 = 2; 2+2=4; 4+5=9 ș.a.

Aceeași regulă se folosește și pentru completarea restului de coloane: (2), (3) și (4).

Însumînd acum cifrele, pe coloane, situate mai sus de linia centrală, obținem sumele ajutătoare $a_i (i = 1, 2, 3, 4,...)$; făcînd aceeași însumare

și pentru cifrele situate mai jos de linia centrală obținem sumele ajutătoare b_i *). Adunînd, pe coloane, sumele ajutătoare a_i și b_i obținem sumele S_i; prin diferență rezultă d_i. Pentru exemplul din tabelul 2.2 avem:

$$S_1 = 170 + 281 = 451 \qquad d_1 = 170 - 281 = -111$$
$$S_2 = 233 + 340 = 573 \qquad d_2 = 233 - 340 = -107$$

Datele obținute fac posibil calculul momentelor obișnuite (tabelul 2.2).

Pentru verificarea rezultatelor obținute vom aplica aceeași modalitate folosită și în cazul procedeului direct. În acest scop calculăm, în penultima coloană a tabelului 2.2, abaterile $\frac{(x_i - x_0)}{h} + 1$ însemnate prin $\xi + 1$. Astfel, în dreptul bazei arbitrare ($x_0 = 7$) înscriem abaterea $+1$ deoarece:

$$\frac{7-7}{1} + 1 = +1$$

În funcție de noile abateri obținute și de frecvențele n_i, folosind tabelul din anexa 1, completăm ultima coloană a tabelului 2.2. În raport cu suma rezultată, după formula (2.9), obținem:

$$m_4^* = \frac{39\,955}{192} = 208{,}1$$

$$m_0 + 4\,m_1 + 6\,m_2 + 4\,m_3 + m_4 = 208{,}1$$

Calculele fiind corecte, vom stabili acum momentele centrale, pe baza formulelor cunoscute:

$$\mu_2 = m_2 - m_1^2 = 7{,}986$$
$$\mu_3 = m_3 - 3\,m_2\,m_1 + 2\,m_1^3 = 10{,}085$$
$$\mu_4 = m_4 - 4\,m_3\,m_1 + 6\,m_2\,m_1^2 - 3\,m_1^4 = 182{,}519$$

Despre corectitudinea calculelor efectuate ne putem convinge aplicînd formulele (2.10) și (2.11).

Momentele ajustate rezultă aplicînd formulele SHEPPARD:

$$\mu_2' = \mu_2 - 0{,}083 = 7{,}986 - 0{,}083 = 7{,}903$$
$$\mu_3' = \mu_3 = 10{,}085$$
$$\mu_4' = \mu_4 - 0{,}5\,\mu_2 + 0{,}029 = 178{,}555$$

Din cele prezentate rezultă că procedeul sumelor este cu mult mai simplu decît procedeul descris anterior. Totuși, în situația prelucrării unui număr mare de serii statistice, calculul manual necesită mult timp, ceea ce frînează uneori valorificarea integrală a materialului recoltat. În noile condiții, cînd s-au pus în funcțiune calculatoare electronice, capabile să efectueze mii și zeci de mii de operații pe secundă, dificultatea semnalată se înlătură.

---

* Cînd seria x_i se înscrie în ordine crescătoare de sus în jos, mărimile ajutătoare a_i se obțin prin însumarea cifrelor situate mai jos de linia centrală, iar b_i — prin însumarea celor situate mai sus de linia centrală (tab. 3.9).

2.3. MEDIA ARITMETICĂ, MEDIA PĂTRATICĂ ŞI MEDIA GEOMETRICĂ

Media aritmetică $\bar{x}$ este parametrul cel mai des folosit în statistică, Pentru valori observate negrupate media se calculează astfel:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.22}$$

unde:

$x_1, x_2, \ldots, x_n$ sînt valori observate;
n — numărul valorilor observate.

Pentru valori grupate, cunoscînd deci frecvenţele pentru fiecare clasă, se foloseşte formula *mediei ponderate*:

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \ldots + x_n n_n}{n_1 + n_2 + n_3 + \ldots + n_n} = \frac{\sum_{i=1}^{n} x_i n_i}{\Sigma n_i} \tag{2.23}$$

În cazul valorilor grupate, pentru calcularea mediei se foloseşte şi procedeul momentelor. Acesta se bazează pe formula:

$$\bar{x} = x_o + m_1 h \tag{2.24}$$

În legătură cu folosirea mediei aritmetice este de menţionat faptul — deosebit de important — că mediile generale calculate pentru colectivităţi eterogene, nestratificate, ascund deosebirile calitative dintre diferitele grupe de unităţi omogene luate în studiu. De aceea, utilizarea mediilor generale a fost criticată. Apar astfel nestabile şi nesigure cifrele privind: productivitatea medie a unor specii forestiere, fără o diferenţiere a arboretelor în raport cu provenienţa lor (sămînţă, lăstari, naturală, artificială etc.); proprietăţile fizico-mecanice ale lemnului prezentate prin medii generale fără o stratificare corespunzătoare a materialului în raport cu condiţiile naturale de creştere; procentul mediu de împădurire nediferenţiat pe zone naturalistice sau raioane silvoeconomice; indicele mediu de utilizare industrială a masei lemnoase pe întreprinderi, îndeosebi cînd parchetele nu sînt omogene sub raportul speciilor şi calităţii masei lemnoase primită spre exploatare; procentul mediu de reuşită a împăduririlor, fără o stratificare prealabilă a suprafeţelor în funcţie de condiţiile de mediu importante pentru reuşita împăduririlor etc.

Media pătratică. Media pătratică se utilizează atunci cînd se urmăreşte atenuarea ponderii unor categorii de variante cu valori mai mici.

Pentru variante negrupate, media pătratică se calculează după formula:

$$\bar{x}_p = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} = \sqrt{\frac{\sum x_i^2}{n}} \tag{2.25}$$

Scanned with OKEN Scanner

Media pătratică ponderată, aplicabilă distribuțiilor statistice, se calculează astfel:

$$\bar{x}_p = \sqrt{\frac{x_1^2 n_1 + x_2^2 n_2 + \ldots + x_n^2 n_n}{n_1 + n_2 + \ldots + n_n}} = \sqrt{\frac{\sum x_i^2 n_i}{\sum n_i}} \tag{2.26}$$

Un exemplu tipic de aplicare a medie pătratice în silvicultură este calculul diametrului mediu al arboretului în raport cu suprafața de bază.

Media geometrică. Pentru funcția determinantă de forma:

$$x_1 \cdot x_2 \cdot x_3 \ldots . x_n = \prod_{i=1}^{n} (x_i) \tag{2.27}$$

corespunde media geometrică:

$$x_g = \sqrt[n]{\Pi (x_i)} \tag{2.28}$$

Pentru serii statistice cu valori grupate avem:

$$\bar{x}_g = \sqrt[\Sigma n_i]{x_1^{n_1} \cdot x_2^{n_2} \cdot x_3^{n_3} \ldots} \tag{2.29}$$

Calculul se face cu ajutorul logaritmilor.

Media geometrică își găsește utilizarea în primul rînd în cazul seriilor dinamice, pentru calculul ritmului mediu de dezvoltare a diferitelor fenomene.

2.4. MEDIANA, QUARTILE, DECILE, MODUL

Mediana este valoarea x a seriei statistice care — atunci cînd toate unitățile seriei sînt aranjate în ordinea de mărime a variantelor — împarte această serie în două grupe egale ca număr. Pentru o distribuție discretă și uniformă cu $2n + 1$ unități, mediana este reprezentată de varianta cu rangul $n + 1$; pentru o distribuție cu $2n$ unități se calculează media aritmetică a variantelor de rangul n și $n + 1$.

Quartilul poate fi: inferior și superior. Primul, Q_I, corespunde valorii care separă primii 25% din termeni; al doilea Q_{III} reprezintă valoarea care separă ultimii 25% din termeni.

În mod similar se definesc *decilele*. Acestea — într-o serie — reprezintă valorile variabilei care împarte volumul colectivității în 10 părți egale.

Mediana, quartilele și decilele reprezintă cazuri particulare ale *quantilelor*. Quantilul se poate defini ca fiind variabila căreia îi corespunde o frecvență relativă cumulată dată. Astfel, mediana este valoarea căreia îi corespunde o frecvență relativă cumulată egală cu 0,5.

Quantilele se pot ușor determina pe cale grafică. În acest scop se construiește *ogiva lui Galton*, luînd în sistemul coordonatelor rectangulare valorile caracteristicii ca abscise și ale frecvențelor cumulate (absolute sau

Scanned with OKEN Scanner

relative) ca ordonate (fig. 2.2). Mediana are o semnificație geometrică simplă: ordonata corespunzătoare medianei împarte suprafața poligonului de frecvență în două părți egale. Astfel, ducînd o paralelă la axa absciselor prin ordonata corespunzătoare frecvenței cumulate totale împărțită cu 2, tăiem ogiva într-un punct a cărui abscisă este mediana seriei statistice luate în studiu. În mod analog se determină și alte quantile.

Mediana se poate determina și după formula:

$$M_e = X_{Me} + \frac{h\left(\frac{\sum n_i}{2} - S_n\right)}{n_{Me}} \qquad (2.30)$$

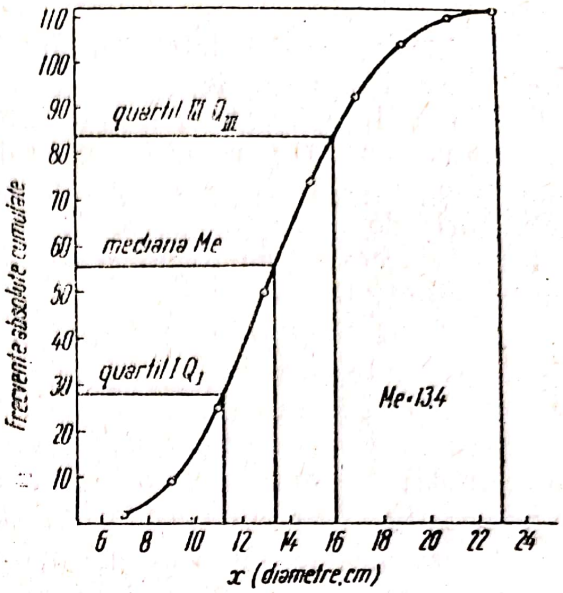


Fig. 2.2. Determinarea quantilelor pe cale grafică

în care:

X_{Me} este limita inferioară a intervalului în care cade mediana;
h — mărime intervalului;
Σn_i — frecvența cumulată totală;
S_n — frecvența cumulată pînă la intervalul median;
n_{Me} — frecvența corespunzătoare intervalului median.

Cu datele, reprezentînd distribuția arborilor pe categorii de diametre într-un arboret de molid, se exemplifică modul de calcul al medianei după formula (2.30) (tabelul 2.3).

Tabelul 2.3

Calculul medianei

Intervale de grupare, cm	Mijlocul intervalului	Frecvența absolută	Frecvența absolută cumulată
5— 7	6	2	2
7— 9	8	7	9
9—11	10	17	26
11—13	12	25	51
13—15	14	23	74
15—17	16	18	92
17—19	18	12	104
19—21	20	6	110
21—23	22	2	112
Total	—	112	—

$$\frac{\Sigma n_i}{2} = \frac{112}{2} = 56;\ X_{Me} = 13;\ h = 2;\ S_n = 51;\ n_{Me} = 23$$

$$M_e = 13 + \frac{2(56 - 51)}{23} = 13{,}4 \text{ cm.}$$

Scanned with OKEN Scanner

Mediana, ca și celelalte quantile, prezintă o proprietate deosebit de folositoare și anume: depinde într-o măsură mai mică decît media aritmetică de forma distribuției, fiind deci mai stabilă. Datorită acestei proprietăți mediana poate fi utilizată pentru distribuții statistice cu un grad mai redus de omogenitate. În asemenea situații, cum sînt repartițiile arborilor pe categorii de diametre în arboretele pluriene, mediana nu este influențată mult de frecvențele de mărimi excepționale de la extremitatea inferioară a distribuției.

Modulul (modul, dominanta) se definește ca valoarea variabilei căreia îi corespunde cea mai mare frecvență, adică reprezintă valoarea variabilei cea mai probabilă (fig. 2.3). Distribuțiile omogene, unde unitățile statistice aparțin aceleiași colectivități, au de regulă un singur modul. Asemenea distribuții se numesc *unimodale*. Distribuțiile *plurimodale* (cu mai multe vîrfuri, fig. 2.4) sînt de regulă eterogene.

În practică se obțin uneori distribuții unimodale și în cazul amestecului unor serii de natură diferită, mai ales atunci cînd modulurile acestora au valori apropiate. De aceea, obținerea unei distribuții unimodale nu este un criteriu sigur al omogenității colectivității luată în studiu.

Modulul distribuțiilor discrete se determină citind varianta cu cea mai mare frecvență (absolută sau relativă). În cazul unei distribuții continue, modulul este abscisa corespunzătoare vîrfului curbei de distribuție. Aceasta poate fi calculată după următoarea formulă:

$$M_o = X_{Mo} + \frac{h(n_1 - n_2)}{n_1 - 2n_2 + n_3} \qquad (2.31)$$

în care:

X_{M_o} este valoarea de la limita inferioară a intervalului modal;

n_2; n_1; n_3 — frecvențele corespuzătoare intervalului modal, celui precedent și celui următor.

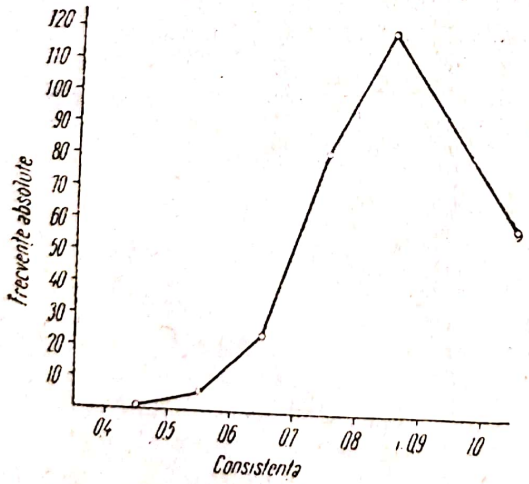


Fig. 2.3. Distribuție unimodală
Distribuția arboretelor de molid în raport cu consistența (clasa a III-a de vîrstă) încadrate în tipul de pădure *Molidiș normal cu Oxalis acetosella*

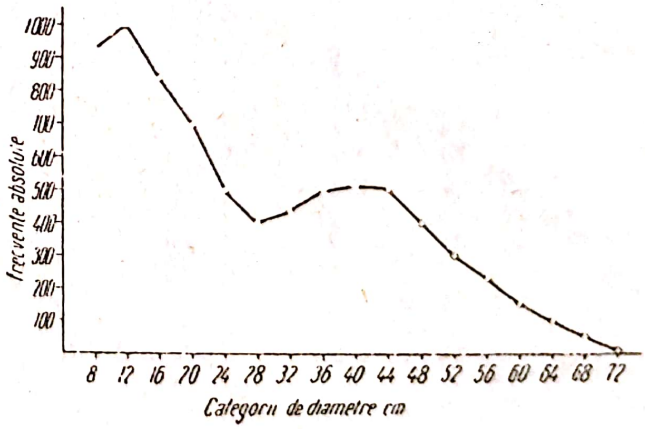


Fig. 2.4. Distribuție bimodală.
Distribuția arborilor pe categorii de diametre într-un arboret bietajat (arbori din două generații)

Scanned with OKEN Scanner

Tabelul 2.4

Consistența arboretelor de molid din clasa a III-a de vîrstă încadrate în tipul de pădure „*Molidiș normal cu Oxalis acetosella*". (Sondaje efectuate în regiunea Suceava)

Consistența (limitele intervalelor)	Mijlocul intervalelor	Frecvențe absolute	Observație
0,40—0,50	0,45	1	Consistența este exprimată pin indicele de densitate
0,50—0,60	0,55	6	
0,60—0,70	0,65	24	
0,70—0,80	0,75	81	
0,80—0,90	0,85	119	
0,90—1,00	0,95	108	
1,00—1,10	1,05	59	
T o t a l	—	398	

Pentru exemplul din tabelul 2.4 avem $h = 0{,}1$; $n_1 = 81$; $n_2 = 119$; $n_3 = 108$; $x_{Mo} = 0{,}8$. Prin urmare:

$$M_o = 0{,}80 + \frac{0{,}1\ (81 - 119)}{81 - 2 \times 119 + 108} = 0{,}88$$

Modulul poate da indicații prețioase în cercetarea statistică forestieră, mai ales atunci cînd prezintă interes valoarea cea mai frecventă, ca, de exemplu:

— vîrsta arboretelor la care doborîturile de vînt sînt mai numeroase; diametrul arborilor căruia îi corespunde cea mai mare frecvență a lemnului pentru derulaj; vîrsta arboretelor la care anumite defecte ale lemnului sînt mai numeroase; consistența caracteristică a arboretelor (cea mai frecvent întîlnită) pentru diferite tipuri de pădure etc.

P e a r s o n stabilește o relație între $\bar{x}$, M_e și M_o, utilă pentru unele aplicații practice:

$$M_o = \bar{x} - 3\ (\bar{x} - M_e) \qquad (2.32)$$

În cazul unei distribuții perfect simetrice

$$\bar{x} = M_o = M_e$$

Pentru distribuțiile asimetrice aceste valori au abateri una față de alta, astfel încît diferențele dintre ele constituie o măsură pentru sesizarea asimetriei unei distribuții.

2.5. INDICATORII VARIABILITĂȚII

Valoarea medie, indicînd doar centrul de grupare a variantelor, nu oferă decît o primă imagine asupra caracterului distribuției studiate, fără să spună nimic asupra gradului de variabilitate. Spre exemplu, caracterizarea proprietăților fizico-mecanice și chimice ale lemnului la speciile forestiere

Scanned with OKEN Scanner

prin cifre medii, deşi utilă pentru o primă cunoaştere a speciei, este insuficientă din punct de vedere ştiinţific şi practic, deoarece o variaţie prea mare a variantelor face calculele statistice şi economice nesigure.

Tot atît de insuficientă este cifra privind consistenţa medie a arboretelor încadrate într-un anumit tip de pădure, sau mediile privind caracteristicile biometrice ale semințelor forestiere. Pentru plinătatea caracterizării distribuţiilor statistice este necesară cunoaşterea modului cum variantele se repartizează în jurul mediei; se impune o caracterizare cantitativă a variabilităţii, „a organizării lăuntrice a distribuţiei"; interesează gradul de variaţie a variantelor în jurul centrului de grupare. Pentru măsurarea acestei împrăştieri se folosesc mai mulţi indicatori, dintre care mai importanţi sînt următorii: amplitudinea variaţiei, dispersia, abaterea medie pătratică, coeficientul de variaţie.

Amplitudinea variaţiei (W) reprezintă diferenţa dintre varianta maximă şi varianta minimă a seriei. Acest indicator este puţin stabil şi depinde într-o măsură prea mare de numărul de măsurători efectuate.

Dispersia (s^2). O măsură mai bună a variaţiei valorilor caracteristicii este media abaterilor faţă de medie, ridicate la pătrat, cunoscută sub denumirea de *dispersie*. Expresia dispersiei este următoarea:

$$s_2^2 = \frac{\Sigma (x_i - \bar{x})^2 n_i}{\Sigma n_i} \tag{2.33}$$

Extrăgînd rădăcina pătrată din valoarea dispersiei obţinem *abaterea medie pătratică* (s)

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2 n_i}{\Sigma n_i}} \tag{2.34}$$

Pentru serii statistice de volum mic se utilizează de obicei formulele:

$$s^2 = \frac{\Sigma (x_i - \bar{x})^2 n_i}{\Sigma n_i - 1} \tag{2.35}$$

$$s = \sqrt{\frac{\Sigma (x_i - \bar{x})^2 n_i}{\Sigma n_i - 1}} \tag{2.36}$$

unde, din considerente teoretice, Σn_i este înlocuită prin $\Sigma n_i - 1$.

Calculul direct al dispersiei şi al abaterii medii pătratice nu necesită explicaţii speciale. Prezintă însă dezavantaje din punct de vedere al volumului de lucrări. Desfăşurarea calculelor se poate urmări în tabelul 2.5, unde se prezintă distribuţia a 200 puieţi de plop alb în raport cu înălţimea lor.

Cu mult mai simplu este *procedeul momentelor*. La baza acestui procedeu stau relaţiile:

$$s = h\sqrt{\mu_2} \quad ; \quad s^2 = h^2 \mu_2$$

sau:

$$s^2 = h^2 (m_2 - m_1^2) \tag{2.37}$$

$$s = h\sqrt{m_2 - m_1^2} \tag{2.38}$$

Scanned with OKEN Scanner

Tabelul 2.5

Determinarea dispersiei și a abaterii medii pătratice prin procedeul direct (Distribuția a 200 puieți de plop alb în raport cu înălțimea lor)

Înălțimea puieților, în cm x_i	Numărul puieților n_i	$x_i\, n_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$
20	4	80	—58,20	3 387,24	13 548,96
30	8	240	—48,20	2 323,24	18 585,92
40	11	440	—38,20	1 459,24	15 051,64
.					.
.					.
.					.
.					.
Σ	200	15 640	—	—	121 412,96

$$\bar{x} = \frac{\Sigma\, x_i\, n_i}{\Sigma\, n_i} = \frac{15\,640}{200} = 78{,}20$$

$$s^2 = \frac{\Sigma\,(x_i - \bar{x})^2\, n_i}{\Sigma n_i} = \frac{121\,412{,}96}{200} = 610{,}11$$

$$s = \sqrt{610{,}11} = 24{,}70$$

După cum s-a arătat, momentele obișnuite m_i se pot calcula fie prin procedeul direct, fie prin procedeul sumelor.

În tabelul 2.6 se prezintă desfășurarea calculelor pentru determinarea dispersiei și a abaterii medii pătratice prin procedeul momentelor calculate direct. Mersul calculelor este identic cu cel prezentat în tabelul 2.1, cu excepția că pentru determinarea valorilor s^2 și s este suficientă cunoașterea primelor două momente m_1 și m_2.

Calculul momentelor prin procedeul sumelor este cu mult mai comod, deoarece toate înmulțirile sînt înlocuite prin simple adunări. Desfășurarea calculelor se poate urmări în tabelul 2.7.

În cercetările din economia forestieră se întîlnesc frecvente cazuri cînd rezultă mai multe serii statistice pentru aceeași caracteristică. Asemenea distribuții, datorită avantajelor oferite de procedeul sumelor, pot fi prelucrate în comun, rezultînd astfel însemnate economii de timp. Totodată se asigură controlul pe parcurs al efectuării calculelor.

Ca exemplu putem da o colectivitate formată din 236 arbori dintr-un arboret echien de pin, în vîrstă de 70 ani, cărora li s-au măsurat creșterile pe rază la înălțimea pieptului din patru părți (sud, est, nord și vest), în scopul evidențierii efectului expoziției asupra creșterii în diametru. După gruparea datelor inițiale în clase, au rezultat patru serii statistice ale aceleiași caracteristici (creșterea pe rază). Însumînd, obținem a cincea serie care cuprinde totalul de 944 măsurători efectuate la cei 236 arbori.

Scanned with OKEN Scanner

Tabelul 2.6

Determinarea dispersiei și a abaterii medii pătratice prin metoda momentelor (varianta I). Creșterea în diametru pe *5* ani la arbori într-un arboret echien de molid de *50* ani (vezi tabelul *1.2*)

Creșterea în mm x_i	Frecvențe n_i	$\xi_i = \frac{x_i - x_0}{h}$	$\xi_i\, n_i$	$\xi_i^2\, n_i$
1	2	—6	—12	72
2	12	—5	—60	300
3	17	—4	—68	272
4	19	—3	—57	171
5	31	—2	—62	124
6	22	—1	—22	22
$x_0 = 7$	24	0	0	0
8	23	1	23	23
9	13	2	26	52
10	12	3	36	108
11	8	4	32	128
12	5	5	25	125
13	2	6	12	72
14	0	7	0	0
15	2	8	16	128
Σ	192	—	—111	1 597

$$m_1 = \frac{\Sigma\, \xi_i\, n_i}{\Sigma\, n_i} = \frac{-111}{192} = -0{,}578; \qquad m_2 = \frac{\Sigma\, \xi_i^2\, n_i}{\Sigma\, n_i} = \frac{1\,597}{192} = 8{,}319$$

$$\bar{x} = x_0 + h m_1 = 6{,}423$$

$$\mu_2 = m_2 - m_1^2 = 8{,}319 - (-0{,}578)^2 = 7{,}98; \; h = 1$$

$$s^2 = h^2\, \mu_2 = 7{,}98; \qquad s = h\sqrt{\mu_2} = 2{,}82$$

Desfășurarea prelucrării în comun a acestor cinci serii se poate urmări în tabelul 2.8. După cum se vede s-au format 5 compartimente, corespunzătoare celor cinci serii, păstrînd aceeași coloană x_i pentru toate seriile. În coloanele n_i se trec frecvențele fiecărei serii. În coloana N_i a compartimentului „Total" se înregistrează sumele variantelor n_i, adică:

$N_i = n_i(\mathrm{I}) + n_i(\mathrm{II}) + n_i(\mathrm{III}) + \ldots$ Aplicînd această formulă la totalurile coloanelor n_i, verificăm dacă nu s-au înregistrat greșeli la gruparea datelor, deoarece suma acestor totaluri trebuie să fie egală cu suma coloanei N_i.

În continuare, alegem o singură bază de pornire pentru toate seriile. În raport cu această bază comună, calculăm momentele obișnuite și în funcție de care stabilim momentele centrale, dispersiile și abaterile medii pătratice, folosind în acest scop formulele cunoscute. Verificarea pe parcurs a calculelor se efectuează după cum urmează:

$$a_{\mathrm{I}}^{(1)} + a_{\mathrm{II}}^{(1)} + a_{\mathrm{III}}^{(1)} + \ldots = a_{\mathrm{Total}}^{(1)}$$

$$a_{\mathrm{I}}^{(2)} + a_{\mathrm{II}}^{(2)} + a_{\mathrm{III}}^{(2)} + \ldots = a_{\mathrm{Total}}^{(2)}$$

Scanned with OKEN Scanner

Tabelul 2.7

Determinarea dispersiei şi a abaterii medii pătratice prin metoda momentelor (varianta a II-a). Creşterea în înălţime la *199* arbori de pin, clasa a IV-a vîrstă *

Creşterea în înălţime cm x_i	Numărul de arbori n_i	(1)	(2)
300	5	5	5
280	8	13	18
260	20	33	51
240	29	62	113
220	41	103	—
$x_0 =$ 200	44	—	—
180	25	52	—
160	12	27	58
140	6	15	31
120	4	9	16
100	3	5	7
80	2	2	2
Σn_i	199	—	—
a		216	187
b		110	114
S		326	301
d		106	—

$$m_1 = \frac{d}{\Sigma n_i} = \frac{106}{199} = 0{,}533; \qquad h = 20$$

$$\bar{x} = x_0 + hm_1 = 200 + 20 \times 0{,}533 = 210{,}660;$$

$$m_2 = \frac{S_1 + 2S_2}{\Sigma n_i} = \frac{326 + 602}{199} = 4{,}663; \qquad \mu_2 = m_2 - m_1^2 = 4{,}379;$$

$$s^2 = h^2\ \mu_2 = 1751{,}600; \qquad s = h\sqrt{\mu_2} = 41{,}840$$

$$b_{\mathrm{I}}^{(1)} + b_{\mathrm{II}}^{(1)} + b_{\mathrm{III}}^{(1)} + \ldots = b_{\mathrm{Total}}^{(1)}$$

.

.

.

$$d_{\mathrm{I}} + d_{\mathrm{II}} + d_{\mathrm{III}} \ldots = d_{\mathrm{Total}}$$

Adică:

$$69 + 126 + 113 + 67 = 375$$
$$62 + 94 + 75 + 51 = 282$$

* Date extrase din lucrarea „Determinarea creşterii curente la arborete" (Giurgiu, 1957)

Scanned with OKEN Scanner

$$333 + 250 + 282 + 346 = 1211$$

.

.

.

$$(-264) + (-124) + (-169) + (-279) = -836$$

Prin urmare, calculul acestor sume este corect și se poate trece la determinarea momentelor și a parametrilor distribuțiilor statistice.

Procedeul prezentat este foarte comod, simplu și totodată — prin posibilitatea efectuării controlului calculelor pe parcurs — imprimă rezultatelor obținute un înalt grad de certitudine. Pentru lucrări de mare serie se recomandă folosirea calculatoarelor electronice.

Tabelul 2.8

Prelucrarea în comun a mai multor serii ale aceleiași caracteristici. (Creșterea arborilor pe rază, măsurată din partea sudică — seria I; din partea de est — seria a II-a; din partea de nord — seria a III-a; din partea de vest — seria a IV-a) Pin *70* ani *

Creșterea pe rază, în mm x_i	Seria I			Seria II			Seria III			Seria IV			Total V		
	n_i	(1)	(2)	n_i	(1)	(2)	n_i	(1)	(2)	n_i	(1)	(2)	N_i	(1)	(2)
29	6	6	6	8	8	8	5	5	5	5	5	5	24	24	24
26	6	12	18	10	18	26	9	14	19	5	10	15	30	54	78
23	8	20	38	16	34	60	18	32	51	6	16	31	48	102	180
20	11	31	—	32	66	—	30	62	—	20	36	—	93	195	—
x_0 17	39	—	—	44	—	—	34	—	—	38	—	—	155	—	—
14	62	166	—	54	126	—	54	140	—	50	162	—	220	594	—
11	53	104	167	32	72	124	42	86	142	52	112	184	179	374	617
8	39	51	63	28	40	52	32	44	56	48	60	72	147	195	243
5	12	12	12	12	12	12	12	12	12	12	12	12	48	48	48
Σ	236			236			236			236			944		
a		69	62		126	94		113	75		67	51		375	282
b		333	242		250	188		282	210		346	288		1211	908
S		402	304		376	282		395	285		413	319		1586	1190
d		—264	—		—124	—		—169	—		—279	—		—836	—
m_1		—1,119			—0,525			—0,716			—1,182			—0,885	
m_2		4,279			3,983			4,089			4,453			4,201	
μ_2		3,027			3,707			3,576			3,056			3,418	
$\bar{x}$		13,643			15,425			14,852			13,454			14,348	
s^2		27,243			33,363			32,184			27,504			30,762	
s		5,22			5,78			5,67			5,24			5,55	

* Extras din lucrarea „Determinarea creșterii curente la arborete" (Giurgiu, 1957)

Scanned with OKEN Scanner

La întocmirea programului de calcul se dovedește mai corespunzătoare formula:

$$s^2 = \frac{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}{n-1} \qquad (2.39)$$

care reprezintă o transformare a formulei de bază (2.36).

Formula (2.39) este utilă și pentru calcule efectuate prin procedee clasice, mai ales pentru serii negrupate și formate dintr-un număr redus de variante. Calculele se fac ușor în posesia unei mașini electrice de calculat (tabelul 2.9).

Tabelul 2.9

Calculul abaterii medii pătratice după formula (2.39). (Inventarieri prin suprafețe de probă circulare într-un arboret plurien; Ocolul silvic Sinaia)

Suprafața de bază x_i	x_i^2
149	22 201
77	5 929
109	11 881
.	.
.	.
.	.
$\sum x_i = 5\,426$	$\sum x_i^2 = 754\,405$

$$n = 43$$

$$\bar{x} = \frac{5\,426}{43} = 126$$

$$s^2 = \frac{754\,405 - \frac{5\,426^2}{43}}{42} = 1\,659; \quad s = 40{,}7$$

$$s\% = \frac{40{,}7}{126}\,100 = 32{,}3$$

Un alt parametru al distribuțiilor statistice este *coeficientul de variație* $s\%$*. Prin definiție, acesta reprezintă abaterea medie pătratică exprimată în procente față de media aritmetică, adică

$$s\% = \frac{s}{\bar{x}}\,100 \qquad (2.40)$$

Pentru exemplul din tabelul 2.9 avem:

$$\bar{x} = 126; \; s = 40{,}7; \; s\% = 32\%.$$

* In diferite publicații pentru coeficientul de variație se folosesc simbolurile: c_v, V, V_σ.

Folosirea coeficientului de variație este deosebit de convenabilă, permițînd compararea mărimilor abaterii medii pătratice la mai multe colectivități, mai ales la colectivitățile caracterizate prin însușiri diferite.

Mărimea coeficientului de variație este legată de gradul de omogenitate a colectivității statistice din punctul de vedere al caracteristicii luate în studiu. Pentru colectivități mai puțin omogene $s\%$ poate depăși 30%, ceea ce sugerează ideea unor stratificări corespunzătoare a unităților statistice în vederea obținerii unor distribuții mai omogene.

Cunoașterea coeficientului de variație prezintă un mare interes practic și științific. Pentru multe caracteristici, în anumite condiții, se pot stabili valori medii suficient de stabile. Pentru arboretele echiene s-au scos în evidență valorile din tabelul 2.10.

Tabelul 2.10

Coeficienți de variație pentru arboretele echiene

Caracteristici	Coeficientul de variație, %	
	In cadrul arboretului total	In cadrul categoriilor de diametre centrale
Diametre	20—30	—
Înălțimi	10—20	6—10
Indici de formă obișnuiți	6—10	5— 8
Indici de formă reali	5— 6	4— 5
Coeficienți de formă obișnuiți	8—12	6—10
Coeficienți de formă reali	6— 7	5— 7
Grosimea cojii la înălțimea pieptului	25—35	15—20
Diametrul coroanelor	25—35	15—20
Înălțimea elagată	8—12	6— 8
Creșterea în înălțime	20—30	18—20
Creșterea în diametru	35—60	25—35
Creșterea în suprafața de bază	60—90	25—40
Creșterea în volum	60—90	25—40
Procentul creșterii în volum	25—35	25—35

Găsim, de exemplu, că diametrele sînt mai variabile decît înălțimile sau că indicii de formă și coeficienții de formă sînt mai puțin variabili decît înălțimile. O variabilitate accentuată se constată în cazul creșterilor. Rezultă deci că, pentru aceeași colectivitate, variabilitatea diferă de la caracteristică la caracteristică. Pe de altă parte se constată că, prin stratificare, se reduce coeficientul de variație, lucru care nu trebuie pierdut din vedere.

Coeficientul de variație al unuia și aceluiași caracter se modifică în raport cu vîrsta; de obicei se constată o diminuare a coeficientului de variație pe măsură ce arboretele înaintează în vîrstă, fapt explicabil din punct de vedere biologic. Astfel, pentru înălțimi, în primii ani de la plantare, coeficientul de variație este de 30 — 40%, iar la arboretele exploatabile acesta scade pînă la 8 — 10%.

Scanned with OKEN Scanner

III

DISTRIBUȚII STATISTICE FRECVENT ÎNTILNITE IN CERCETĂRILE DIN DOMENIUL SILVICULTURII

3.1. DISTRIBUȚIA BINOMIALĂ

Una dintre cele mai simple distribuții, dar destul de frecvent întîlnită în practică, este distribuția *binomială*. Această distribuție apare la rezolvarea următoarei probleme: se efectuează n încercări succesive *independente*, știind că în fiecare încercare probabilitatea de apariție a evenimentului dat este *constantă* și egală cu p; să se determine probabilitatea apariției evenimentului respectiv de x ori. Asemenea încercări succesive pot fi rezultatele verificării calității pieselor industriale, luate cîte una dintr-o serie de loturi pentru care procentul produselor necorespunzătoare este același. Probabilitatea (frecvența) că variabila întîmplătoare va avea o valoare oarecare x este dată de formula:

$$f_n(x) = \frac{n!}{x!(n-x)!} p^x \, q^{n-x} \tag{3.1}$$

unde x reprezintă numărul elementelor cu caracteristica dată din cele n elemente extrase ($x = 0, 1, 2, 3, \ldots, n$), iar $q = 1 - p$.

Aceasta este funcția de frecvență a distribuției binomiale.

Pentru aplicații practice se aplică următoarea formulă de recurență:

$$f_n(x+1) = f_n(x) \frac{n-x}{x+1} \cdot \frac{p}{q} \tag{3.2}$$

Potrivit acestei formule este suficient să se calculeze probabilitatea $f_n(o)$ corespunzătoare valorii $x = 0$, cu ajutorul formulei (3.1), și apoi pe baza formulei (3.2) să se calculeze celelalte probabilități.

Calculul funcției de frecvență pentru valori mari ale lui n este foarte anevoios din cauza factorialilor; de aceea în practică se folosește formula asimptotică următoare:

$$n! \approx n^n e^{-n} \sqrt{2\pi n} \tag{3.4}$$

Dispersia teoretică este dată de formula:

$$s^2 = n\ p\ q \tag{3.5}$$

Media aritmetică se calculează astfel:

$$\bar{x} = n\ p \tag{3.6}$$

de unde:

$$p = \frac{\bar{x}}{n} \tag{3.7}$$

Se știe că:

$$q = 1 - p \tag{3.8}$$

Spre exemplificare, să considerăm distribuția numărului de plăci aglomerate din lemn, necorespunzătoare standardului, obținute într-un schimb, știind că, în condițiile procesului tehnologic respectiv, probabilitatea ca o piesă să fie necorespunzătoare este egală cu 0,02 . Să se calculeze probabilitatea ca într-o selecție de 100 piese extrase să fie defecte 0, 1, 2, ..., 100 piese.

$$p = 0{,}02;\quad q = 0{,}98;\quad n = 100;\quad x = 0,1,2,\ \ldots,\ 100$$

După formula (3.1) pentru $x = 0$ obținem probabilitatea de 0,1326. Restul probabilităților (tabelul 3.1) s-au calculat după formula de recurență (3.2). Analizînd datele obținute rezultă că probabilitatea evenimentului $x > 5$ (mai mult de 5 piese defecte) este foarte mică, deci în mod practic acest eveniment este imposibil.

Tabelul 3.1

Calculul distribuției binomiale pentru $n = 100$, $p = 0{,}02$, $q = 0{,}98$

Numărul pieselor defecte x	Frecvențe relative	Frecvențe relative cumulate
0	0,1326	0,1326
1	0,2707	0,4033
2	0,2734	0,6767
3	0,1823	0,8590
4	0,0902	0,9492
5	0,0353	0,9845
6	0,0114	0,9959
7	0,0031	0,9990

Funcția de frecvență a distribuției binomiale (formula 3.1) poate fi cu succes folosită în practică pentru ajustarea unor distribuții experimentale. Să luăm, spre exemplu, totalitatea bușteni lor de molid avînd aproximativ aceleași dimensiuni și o proveniență asemănătoare. Din această colectivitate generală s-au extras 60 probe, proba avînd 6 bușteni. Pentru fiecare probă

Scanned with OKEN Scanner

s-a stabilit numărul pieselor cu noduri vicioase. Repartiția probelor în raport cu numărul pieselor cu noduri vicioase este următoarea:

Grafic, distribuția experimentală este prezentată în fig. 3.1. Aceasta se poate ajusta potrivit legii distribuției binomiale, respectiv după formula (3.1).

Tabelul 3.2

Numărul pieselor cu noduri vicioase x	Numărul cazurilor
0	4
1	13
2	21
3	15
4	6
5	1
6	—
Total	60

În baza datelor din tabelul 3.2, stabilim:

$$\bar{x} = 2{,}15 \; ; \quad s^2 = 1{,}28 \; ; \quad s = 1{,}13 \; ; \; n = 6$$

După formula (3.7) avem:

$$p = \frac{\bar{x}}{n} = \frac{2{,}15}{6} = 0{,}358$$

de unde:

$$q = 1 - p = 1 - 0{,}358 = 0{,}642 \; ; \quad \frac{p}{q} = \frac{0{,}358}{0{,}642} = 0{,}55$$

Potrivit formulei (3.1), stabilim frecvența relativă teoretică pentru $x = 0$:

$$f_n(0) = 0{,}642^6 = 0{,}071$$

Acum, după formula de recurență (3.2) obținem restul frecvențelor relative. Astfel:

$$f(1) = 0{,}55 \cdot \frac{6}{1} \cdot 0{,}071 = 0{,}24$$

$$f(2) = 0{,}55 \cdot \frac{5}{2} \cdot 0{,}24 = 0{,}33 \text{ etc.}$$

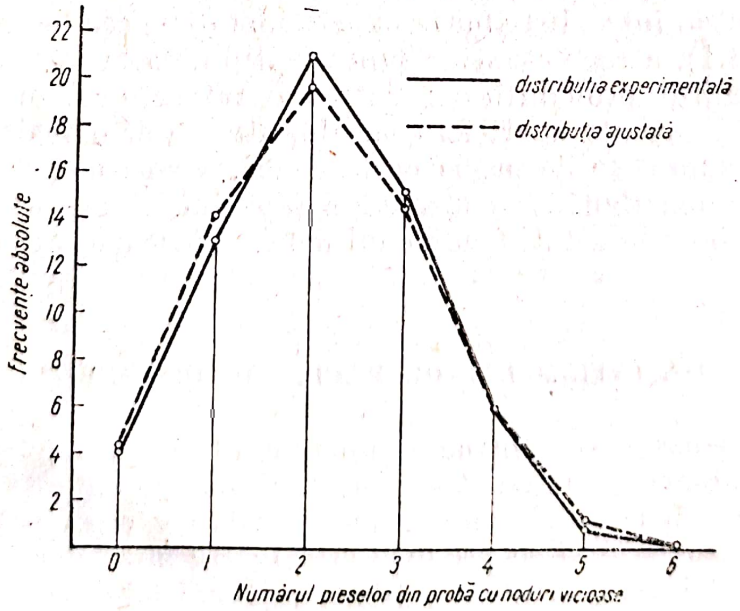


Fig. 3.1. Distribuția unor probe de bușteni în raport cu numărul pieselor cu noduri vicioase

Scanned with OKEN Scanner

Înmulțind frecvențele relative cu numărul total de observații (60) obținem frecvențele teoretice în valori absolute (tabelul 3.3).

Tabelul 3.3

Comparație între frecvențele experimentale și frecvențele ajustate potrivit legii distribuției binomiale (pentru distribuția prezentată în tabelul 3.2)

Numărul pieselor cu noduri vicioase x	Numărul cazurilor	Valori empirice ale funcției de frecvență	Valori teoretice ale funcției de frecvență	Frecvențe absolute ajustate
0	4	0,07	0,07	4
1	13	0,22	0,24	14
2	21	0,35	0,33	20
3	15	0,25	0,24	15
4	6	0,10	0,10	6
5	1	0,01	0,02	1
6	—	0,002	0,002	—
	60	1,00	1,00	60

În situațiile cînd p este mai mare sau mai mic decît 0,5 funcția de frecvență este asimetrică, în special atunci cînd n este mic; o dată cu creșterea lui n asimetria se reduce treptat, astfel încît se apropie din ce în ce mai mult de funcția de frecvență a distribuției normale (§ 3.3).

Legea distribuției binomiale este aplicabilă ori de cîte ori studiem fenomene independente a căror probabilitate de apariție (sau neapariție) este cunoscută și are o valoare constantă. Probabilitatea p poate fi calculată în baza unei selecții anterioare, așa cum s-a procedat în exemplul din tabelul 3.2.

Neconcordanța dintre distribuția experimentală și cea teoretică, calculată după formula (3.1), atrage atenția asupra faptului că la realizarea diverselor frecvențe, în cadrul probabilității date, nu intervine numai întîmplarea. Astfel, studiind frecvența apariției arborilor atacați de o boală criptogamică, în cadrul unor suprafețe de probă cu n arbori, vom constata o discordanță între cele două distribuții, ca urmare a faptului că atacul se răspîndește ușor de la un arbore la altul, fenomenul nefiind deci datorat numai „întîmplării".

3.2. DISTRIBUȚIA EVENIMENTELOR RARE, SAU DISTRIBUȚIA POISSON

Distribuția POISSON se obține atunci cînd intervine un eveniment rar, putînd fi observat numai dacă numărul sondajelor este mare. Cu alte cuvinte, această distribuție corespunde situațiilor cînd probabilitatea ca un eveniment oarecare să se producă este foarte mică dar constantă. Condiția uniformității este esențială pentru aplicarea legii POISSON.

Distribuția POISSON reprezintă deci „legea evenimentelor rare", fiind utilizabilă în cazul valorilor discontinue ale variabilei.

Scanned with OKEN Scanner

Funcția de frecvență a distribuției POISSON este

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} \tag{3.9}$$

unde:

λ — constantă, reprezentînd media aritmetică și unicul parametru al distribuției;
x — numărul elementelor cu caracteristica dată din cele n extrase (numărul de succese în n probe).

Pentru $x = 0$ avem $f(o) = e^{-\lambda}$

Pentru celelalte valori ale lui x se folosește următoarea formulă de recurență:

$$f(x+1) = f(x) \frac{\lambda}{x+1} \tag{3.10}$$

Se demonstrează că pentru această distribuție particulară

$$s^2 = \lambda \tag{3.11}$$

Graficul din fig. 3.2. arată că la un λ mic, distribuția este unilaterală; odată cu creșterea lui λ distribuția devine din ce în ce mai simetrică, pentru ca la $\lambda = 10$ să se transforme într-o distribuție simetrică (normală). Valorile funcției de frecvență se prezintă în tabelul 3.4

Spre exemplificare, prezentăm distribuția unor suprafețe de probă de 500 m² (amplasate mecanic într-un arboret echien și uniform de molid), în raport cu numărul arborilor bifurcați. Așa cum se știe, apariția arborilor bifurcați este un fenomen relativ rar.

Rezultatele măsurătorilor sînt prezentate în tabelul 3.5.

Potrivit datelor prezentate obținem

$\lambda = \frac{0\times62 + 1\times32 + 2\times6 + 3\times1 + 4\times0}{101} = 0{,}5$ arbori bifurcați; pentru $x = 0$, după formula (3.9) sau tabelul 3.4 avem:

$$f(o) = e^{-0,5} = \frac{1}{\sqrt{2{,}718}} = 0{,}607$$

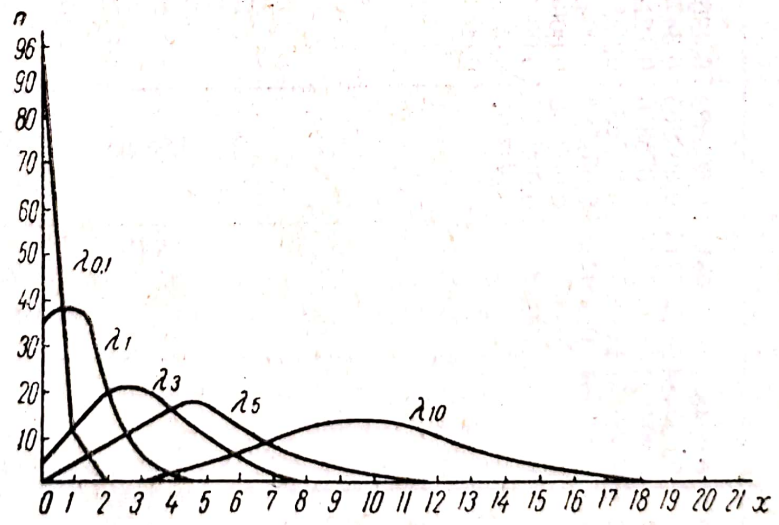


Fig. 3.2. Distribuția POISSON pentru diferite valori ale lui λ

Scanned with OKEN Scanner

Tabelul 3.4

Legea de distribuție a lui Poisson. Funcția de frecvență $f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$

x / λ	0,1	0,3	0,5	0,7	0,9	1	2	4	6	8	10	x
0	0,904837	0,740818	0,606529	0,496585	0,406570	0,367879	0,135335	0,018316	0,002479	0,000335	0,000045	0
1	0,090484	0,222245	0,303265	0,347610	0,365913	0,367879	0,270671	0,073263	0,014873	0,002684	0,000454	1
2	0,004524	0,033337	0,075816	0,121663	0,164661	0,183940	0,270671	0,146525	0,044618	0,010735	0,002270	2
3	0,000151	0,003334	0,012636	0,028388	0,049398	0,061313	0,180447	0,195367	0,089235	0,028626	0,007567	3
4	0,000004	0,000250	0,001580	0,004968	0,011115	0,015528	0,090224	0,195367	0,133853	0,057252	0,018917	4
5		0,000015	0,000158	0,000696	0,002001	0,003066	0,036089	0,156293	0,160623	0,091604	0,037833	5
6		0,000001	0,000013	0,000081	0,000300	0,000511	0,012030	0,104196	0,160623	0,122138	0,063055	6
7			0,000001	0,000008	0,000039	0,000073	0,003437	0,059540	0,137677	0,139587	0,090079	7
8				0,000001	0,000004	0,000009	0,000859	0,029770	0,103258	0,139587	0,112599	8
9						0,000001	0,000191	0,013231	0,068838	0,124077	0,125110	9
10							0,000038	0,005292	0,041303	0,099262	0,125110	10
							0,000007	0,001925	0,022529	0,072190	0,113736	11
							0,000001	0,000642	0,011264	0,048127	0,094780	12
								0,000197	0,005199	0,029616	0,072908	13
								0,000056	0,002228	0,016924	0,052077	14
								0,000015	0,000891	0,009026	0,034718	15
								0,000004	0,000334	0,004513	0,021699	16
								0,000001	0,000118	0,002124	0,012764	17
									0,000039	0,000944	0,007091	18
									0,000012	0,000397	0,003732	19
									0,000004	0,000159	0,001866	20
									0,000001	0,000061	0,000889	21
										0,000022	0,000404	22
										0,000008	0,000176	23
										0,000003	0,000073	24
										0,000001	0,000029	25
											0,000011	26
											0,000004	27
											0,000001	28
											0,000001	29
												30

x — reprezintă numărul apariției fenomenului (numărul de succese în n probe);
λ — numărul mediu al acestor apariții

Scanned with OKEN Scanner

Tabelul 3.5

Repartiția unor suprafețe de probă de *500* m² în raport cu numărul arborilor bifurcați (înregistrări efectuate într-un arboret echien de molid, *70* ani, ocolul silvic Pojorîta)

Numărul arborilor bifurcați x	Numărul suprafețelor de probă	Valorile empirice ale funcției de frecvență	Valorile teoretice ale funcției de frecvență	Frecvențe absolute teoretice
0	62	0,615	0,607	61
1	32	0,316	0,303	31
2	6	0,059	0,076	8
3	1	0,010	0,013	1
4	0	0,000	0,001	0
	101	1,000	1,000	101

Aplicînd acum formula (3.10) sau folosind tabelul 3.4, calculăm restul valorilor teoretice ale funcției de frecvență. Înmulțind aceste frecvențe cu numărul total al suprafețelor de probă, obținem frecvențele absolute ajustate (teoretice). Din tabelul 3.5 rezultă o mare apropiere între valorile empirice și valorile teoretice ale funcției de frecvență, ceea ce arată că fenomenul cercetat urmează legea distribuției POISSON. Ca o concluzie practică se desprinde faptul că în majoritatea loturilor (61%) lipsesc arbori bifurcați; probabilitatea apariției unor loturi cu 4 arbori bifurcați este foarte redusă și anume 0,001.

Legea POISSON este urmată de multe fenomene întîlnite în silvicultură. Poate fi întîlnită la studiul frecvenței anumitor dăunători sau defecte la arbori ori loturi de arbori, la studiul unor anomalii ale lemnului, la studiul frecvenței arborilor pe anumite suprafețe de sondaj etc. Această distribuție a fost identificată și la studiul distanțelor dintre arbori într-un arboret echien. A fost întîlnită și la studiul repartizării teritoriale a unor zoocenoze etc.

3.3 DISTRIBUȚIA NORMALĂ

În activitatea practică și în cercetare se întîlnesc multe caracteristici care variază sub influența unui număr foarte mare de cauze, fiecare dintre acestea acționînd independent una de alta. În asemenea situații aportul fiecărei cauze este foarte mic în raport cu efectul influenței comune a complexului de cauze care acționează asupra caracteristicii respective. Abaterea fiecărei valori de la medie sau de la o anumită valoare prescrisă variază întîmplător de la o unitate la alta. Conform teoremei lui LIAPUNOV aceste abateri urmează legea lui GAUSS, denumită și *legea distribuției normale*.

Reprezentată prin poligonul de frecvență, distribuția normală se prezintă sub forma unui profil de clopot.

Scanned with OKEN Scanner

Legea lui GAUSS, folosită inițial la studiul repartiției erorilor, se exprimă cu ajutorul formulei:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \tag{3.12}$$

în care σ și a sînt parametrii distribuției normale. Primul parametru σ reprezintă abaterea medie pătratică teoretică, stabilită în baza unui mare număr de măsurători; al doilea parametru a reprezintă media aritmetică teoretică a colectivității generale*. Pentru colectivități de selecție, σ și a se estimează cu ajutorul abaterii medii pătratice empirice s și, respectiv, prin intermediul mediei aritmetice empirice $\bar{x}$. Cînd volumul colectivității de sondaj este mare, diferențele dintre $\bar{x}$ și a și respectiv dintre s și σ sînt reduse și practic neglijabile.

Funcția de distribuție a distribuției normale se obține prin integrarea funcției de frecvență (3.12).

Funcția de distribuție arată probabilitatea ca valorile caracteristicii să fie mai mici sau egale cu o valoare dată:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-a)^2}{2\sigma^2}} dx \tag{3.13}$$

În fig. 3.3 se reprezintă funcția de frecvență potrivit formulei (3.12) și funcția de distribuție a distribuției normale, potrivit formulei (3.13).

Suprafața cuprinsă între $a - \sigma$ și $a + \sigma$ reprezintă 68,3% din suprafața totală, respectiv cuprinde majoritatea cazurilor și anume 68,3% din totalul cazurilor. În intervalul $a - 2\sigma$ și $a + 2\sigma$ sînt cuprinse 95,4% din totalul cazurilor. În afara acestor limite se găsesc numai 4,6% din cazuri (tabelul 3.6 coloana 7), iar în afara limitelor $+3\sigma$ și -3σ rămîn numai 0,27% din cazuri.

De aici regula importantă: *variabila întîmplătoare x distribuită după legea normală are abateri mai mici decît $\pm 3\sigma$*. Potrivit acestei reguli se pot exclude valorile dubioase ale variabilei x.

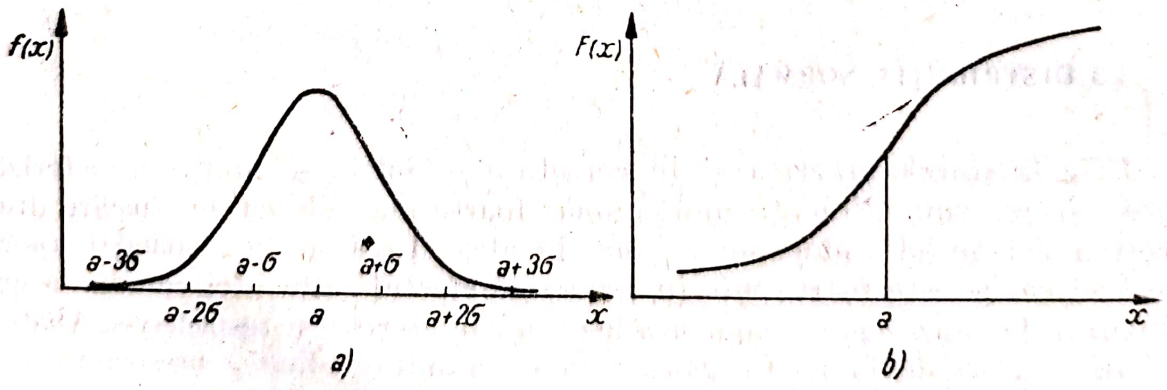


Fig. 3.3. Distribuția normală. Graficul funcției de frecvență (*a*) și graficul funcției de distribuție (*b*)

* Uneori se folosește simbolul μ.

Scanned with OKEN Scanner

Abaterea medie pătratică servește drept măsură a gradului de deschidere a „clopotului" (fig. 3.4).

Pentru a putea compara diferite distribuții normale, s-a introdus indicatorul:

$$u = \frac{x - a}{\sigma} \qquad (3.14)$$

denumit *variabilă normată*. Valorile u exprimă astfel abaterea variabilei față de medie în unități σ.

Avînd în vedere că pentru colectivități de selecție, σ se exprimă prin s, iar a prin x, (3.14) devine:

$$u = \frac{x - \bar{x}}{s} \qquad (3.15)$$

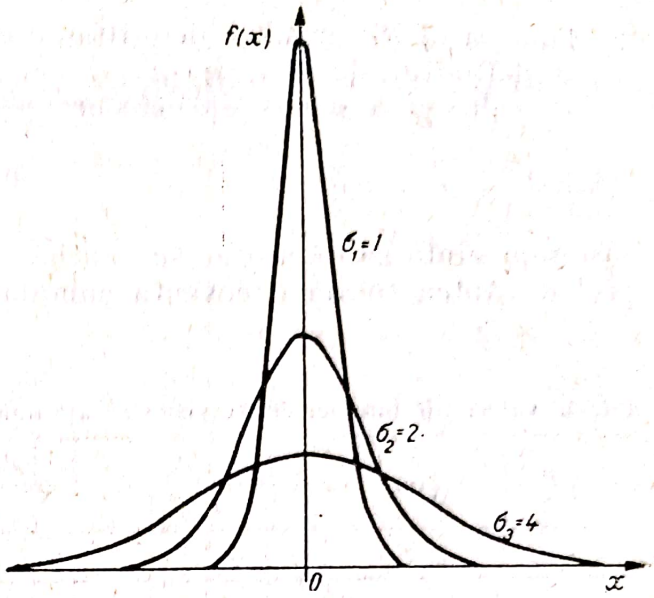


Fig. 3.4. Curbe ale distribuției normale cu parametrii $a = 0$; $\sigma_1 = 1$; $\sigma_2 = 2$ și $\sigma_3 = 3$.

Expresia (3.12) devine o *distribuție normală normată*:

$$f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \qquad (3.16)$$

La fel, (3.13) va avea înfățișarea

$$F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{u^2}{2}} du \qquad (3.17)$$

Pentru distribuția normală normată, media este egală cu 0, iar abaterea medie pătratică este egală cu unitatea (fig. 3.5 *a*).

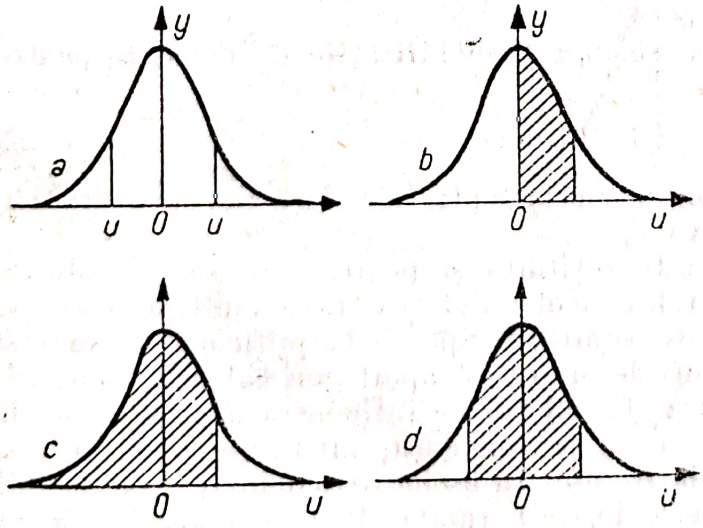


Fig. 3.5. Distribuția normală normată

În tabelul 3.6 sînt date valorile distribuției normale. În coloana 1 sînt date valorile u. Coloana a doua cuprinde valorile $f(u)$ corespunzătoare formulei (3.16). Cifrele din coloana a treia reprezintă valorile funcției LAPLACE (fig. 3.5 *b*):

$$\Phi(u) = \frac{1}{2\pi} \int_{0}^{u} e^{-\frac{u^2}{2}} du \qquad (3.18)$$

Pentru $u = +\infty$, $\Phi(u) = \frac{1}{2}$, iar $2\Phi(u) = 1$

Funcţia (3.18) arată probabilitatea ca variabila să se găsească în intervalul delimitat de $\bar{x}$ şi u.

În coloana a patra sînt trecute valorile

$$\frac{1}{2} + \Phi(u) \tag{3.19}$$

care reprezintă suprafaţa de sub curbă de la $-\infty$ pînă la valoarea u, adică probabilitatea totală (frecvenţa cumulată) pînă la punctul u (fig. 3.5 *c*).

Tabelul 3.6

Cîteva valori ale funcţiei de frecvenţă şi ale funcţiei de distribuţie ale distribuţiei normale

$u = \frac{x - \bar{x}}{s}$	$f(u)$	$\Phi(u)$	$\frac{1}{2} + \Phi(u)$	$1 - \Phi(u)$	$2\Phi(u)$	$1 - 2\Phi(u)$
1	2	3	4	5	6	7
4,0	0,0001	0,49997	0,99997	0,00003	0,99994	0,00006
3,5	0009	49977	99977	00023	99953	00047
3,0	0044	49865	99865	00135	99730	00270
2,5	0175	49379	99379	00621	98758	01242
2,0	0540	47725	97725	02275	95450	04550
1,5	1295	43319	93319	06681	86639	13361
1,0	2420	34135	84135	15865	68269	31731
0,5	3521	19146	69146	30854	38292	61708
0,0	3989	00000	50000	50000	00000	1,00000

În coloana 5 se dau valorile $1 - \Phi(u)$ (3.20)

În coloana a şasea se dau valorile duble din coloana a treia (fig. 3.5 *d*):

$$2\Phi(u) = \int_{-u}^{+u} f(u)\, du \tag{3.21}$$

Această expresie reprezintă probabilitatea totală (sau frecvenţa cumulată) cuprinsă între $-u$ şi $+u$.

În sfîrşit, în ultima coloană se dau probabilităţile de depăşire pentru valoarea u, potrivit formulei

$$q = 1 - 2\Phi(u) \tag{3.22}$$

care arată cu ce probabilitate ne putem aştepta la valori ale caracteristicii în afara intervalului $\pm u$ (anexa 3).

Legea distribuţiei normale este întîlnită şi poate fi cu succes folosită în cercetările forestiere, bineînţeles acolo unde există o multiplicitate de factori independenţi, cu influenţe limitate. Spre exemplificare, luăm distribuţia unui număr de 947 arbori de molid în raport cu valoarea indicelui de formă q_2 (tabelul 3.7). Asupra acestui indice influenţează o mulţime de factori: diametrul, înălţimea, vîrsta, consistenţa, altitudinea, direcţia şi tăria vîntului, modul de gospodărire etc. În asemenea condiţii, este posibil ca distribuţia empirică să urmeze legea normală. Pentru a verifica acest lucru construim curba teoretică, după cum urmează:

Scanned with OKEN Scanner

Tabelul 3.7

Ajustarea distribuției experimentale potrivit legii distribuției normale, după funcția frecvențelor $f(u)$ (Repartiția arborilor de molid din țara noastră în raport cu indicele de formă q_2) (Giurgiu 1955)

Limitele intervalelor	Mediile intervalelor x_i	Frecvențe absolute n_i	Frecvențe relative experimentale	$x_i - \bar{x}$	$u = \frac{x_i - \bar{x}}{s}$	$f(u)$	Frecvențe relative teoretice $\hat{n} = \frac{h}{s} \cdot f(u)$
1	2	3	4	5	6	7	8
0,50—0,52	0,51	7	0,007	—0,156	—2,69	0,0107	0,006
0,53—0,55	0,54	23	0,024	—0,126	—2,17	0,0379	0,020
0,56—0,58	0,57	56	0,059	—0,096	—1,65	0,1023	0,053
0,59—0,61	0,60	96	0,101	—0,066	—1,14	0,2083	0,108
0,62—0,64	0,63	145	0,153	—0,036	—0,62	0,3292	0,170
0,65—0,67	0,66	208	0,221	—0,006	—0,10	0,3970	0,205
0,68—0,70	0,69	174	0,185	0,024	0,41	0,3668	0,190
0,71—0,73	0,72	132	0,139	0,054	0,93	0,2589	0,134
0,74—0,76	0,75	73	0,077	0,084	1,45	0,1394	0,072
0,77—0,79	0,78	24	0,025	0,114	1,97	0,0573	0,030
0,80—0,82	0,81	5	0,005	0,144	2,48	0,0184	0,010
0,83—0,85	0,84	4	0,004	0,174	3,00	0,0044	0,002
Σ		947	1,000				1,000

$\bar{x} = 0{,}666$ $\quad h = 0{,}03$ $\quad s = 0{,}058$ $\quad \frac{h}{s} = \frac{0{,}03}{0{,}058} = 0{,}5172$

În prima coloană a tabelului 3.7. se dau limitele intervalelor, iar în coloana a doua s-au trecut mediile intervalelor. În a treia coloană s-au trecut frecvențele absolute, în funcție de care s-au calculat frecvențele relative (coloana a patra). În continuare s-au calculat diferențele dintre media aritmetică $\bar{x}$ și valorile x_i (coloana a cincea). În baza acestor diferențe,

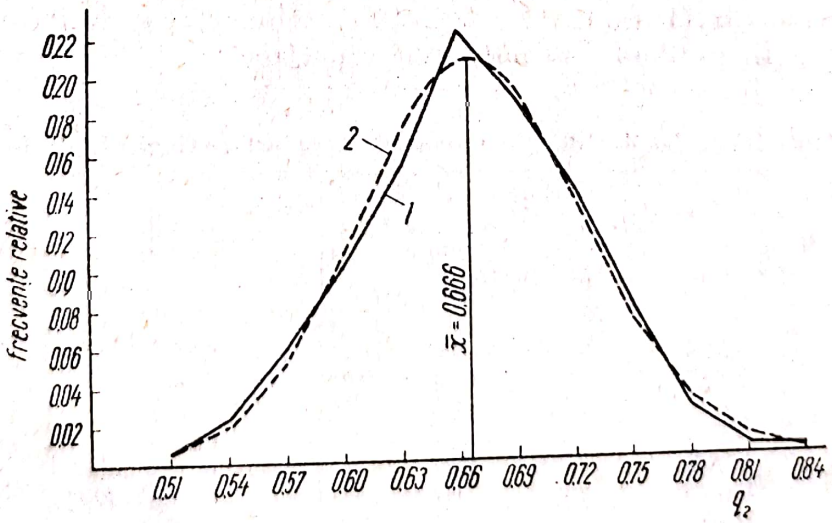


Fig. 3.6. Curba distribuției experimentale (1) și curba distribuției teoretice (2). (Repartiția arborilor de molid din R.S. România în raport cu indicele de formă q_2).

Scanned with OKEN Scanner

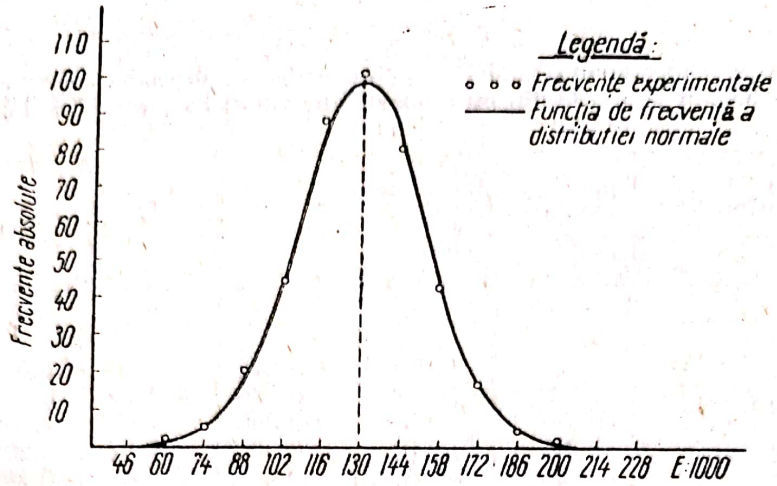


Fig. 3.7. Curba frecvențelor distribuției normale, construită în baza datelor din tabelul 2.1.

după formula (3.15), s-au calculat valorile variabilei normate u. În funcție de u, după formula 3.16 tabelată în anexa 2, s-a calculat distribuția normală normată (coloana 7). De la aceasta se trece la distribuția teoretică, înmulțind valorile din coloana a șaptea cu raportul dintre mărimea intervaului h și abaterea medie pătratică s. În fig. 3.6 sînt reprezentate grafic atît frecvențele experimentale, cît și cele teoretice, ajustate după legea distribuției normale, de unde rezultă o mare apropiere între ele. Aceeași apropiere se constată și în fig. 3.7, construită după datele din tabelul 2.1.

O primă verificare a normalității unei distribuții se poate face prin metoda grafică, folosind *hîrtia probabilistică*. Aceasta este astfel construită încît prin reprezentarea grafică a funcției de distribuție normală se obține o linie dreaptă (fig. 3.8). Pe axa absciselor se trec valorile variabilei x, iar pe axa ordonatelor — frecvențele relative cumulate. Pentru exemplificare, folosim datele studiului indicelui de formă q_2 la pinul silvestru din Carpații Orientali (I a c o v l e v, 1961). Măsurarea indicilor q_2 la 1 020 arbori a scos în evidență următoarea repartiție:

Tabelul 3.8

Repartiția arborilor de pin silvestru în raport cu indicele de formă q_2

q_2		frecvențe absolute	frecvențe relative	frecvențe relative cumulate
media clasei	limita superioară a clasei			
0,50	0,525	12	0,011	0,011
0,55	0,575	44	0,043	0,054
0,60	0,625	140	0,137	0,191
0,65	0,675	384	0,379	0,570
0,70	0,725	372	0,364	0,934
0,75	0,775	64	0,062	0,996
0,80	0,825	4	0,004	1,000
Total	—	1 020	1,000	—

Scanned with OKEN Scanner

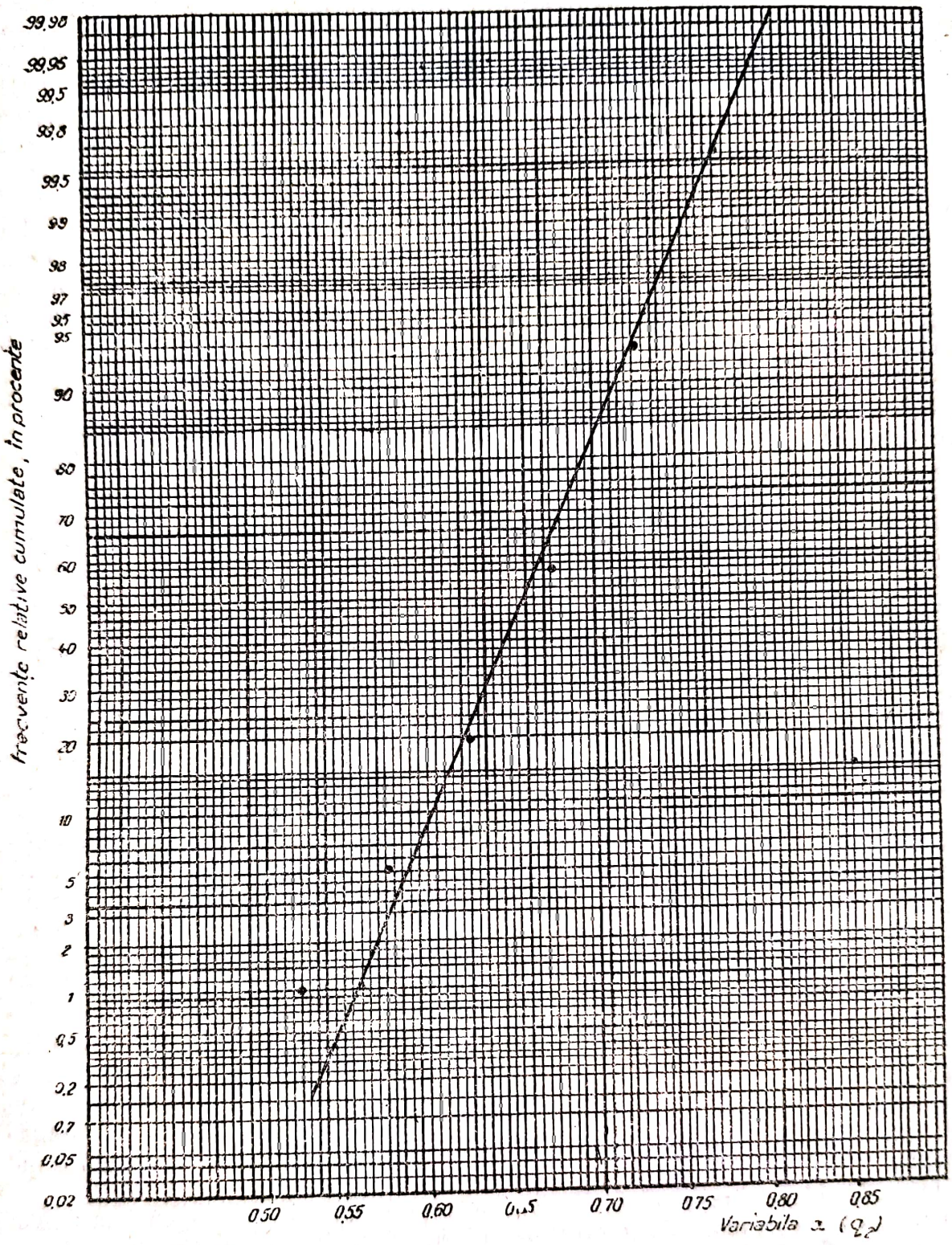


Fig. 3.8. Hîrtia probabilistică (Repartiția arborilor de pin silvestru din Carpații Orientali în raport cu indicele de formă q_2)

Scanned with OKEN Scanner

Frecvențele relative cumulate, pe hîrtia probabilistică, se aranjează de-a lungul unei linii, drepte, ceea ce constituie o primă indicație asupra posibilității aplicării în acest caz a legii distribuției normale (fig. 3.8).

În figura 3.9 se prezintă pe hîrtia probabilistică distribuția arborilor într-un arboret echien de larice în raport cu creșterea curentă în diametru, separat pe categorii de diametre relative (G i u r g i u, 1957). Prin aceasta se scoate în evidență faptul că, în cadrul fiecărei categorii de diametre, repartiția arborilor în raport cu creșterea lor în diametru urmează legea distribuției normale — aspect interesant din punct de vedere biologic. În schimb, în cazul cînd arborii nu se grupează pe categorii de diametre, distribuția arborilor în raport cu creșterea se îndepărtează de curba normală (fig. 3.10), fapt explicabil dacă avem în vedere corelația puternică dintre dimensiunile arborilor și creșterea curentă.

Putem controla dacă o distribuție este normală și prin *metoda probiturilor*. Probiturile se calculează adăugînd o constantă (de obicei 5) la valorile variabilei normate *u*. Pe axa absciselor se trec limitele claselor, iar pe axa ordonatelor se trec valorile probiturilor. Dacă pe acest grafic punctele se ordonează liniar putem afirma că distribuția este normală. Pentru simplificarea calculelor s-au întocmit tabele care dau valorile probiturilor în funcție de frecvențele relative cumulate (*E*. W e b e r, 1957; S t e i n b a c h, 1961).

Probiturile sînt folosite și la examinarea eficacității diferitelor preparate utilizate în fiziologie, toxicologie, entomologie etc. (a n a l i z a p r o b i t) (S n e d e c o r, 1957).

3.4. ASIMETRIA ȘI EXCESUL

În cercetările din domeniul economiei forestiere, ca și în alte domenii se obțin relativ rar curbe de distribuție perfect simetrice, deoarece nu întotdeauna este îndeplinită condiția de independență a multiplicității de factori ce influențează concomitent asupra caracteristicii cercetate. Să luăm, spre exemplu, distribuția puieților în raport cu înălțimea lor într-o plantație tînără, efectuată în condiții staționale omogene. Valorile pentru înălțimi se prezintă sub forma unei curbe de distribuție simetrică, determinată de un complex de cauze variate, cum sînt: factorii ereditari, aprovizionarea cu apă, factorii de nutriție etc. După închiderea stării de masiv situația se schimbă, în sensul că tendința spre lumină face ca un număr tot mai mare de arbori să pătrundă pînă în plafonul superior. Prin aceasta se majorează numărul arborilor cu înălțimi mai mari, respectiv se formează o *asimetrie de dreapta*.

În fig. 3.11 este reprezentată distribuția arborilor în raport cu înălțimea într-un arboret echien de molid de 45 ani (Ocolul silvic Iacobeni, U.P. Orata, u.a.27 a). Asimetria de dreapta este evidentă: modulul este plasat în dreapta mediei.

Pe de altă parte distribuția arborilor în raport cu diametrele se desfășoară în sens opus. În fig. 3.12 se constată o *asimetrie de stînga* determinată

Scanned with OKEN Scanner

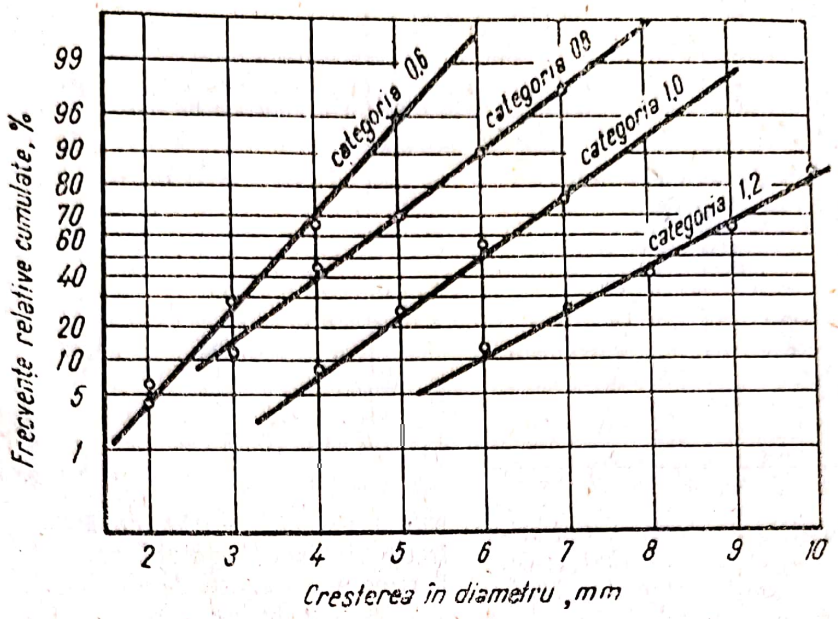


Fig. 3.9. Repartiția arborilor în raport cu creșterea în diametru, separat pe clase de diametre (în valori relative) pentru un arboret echien de larice

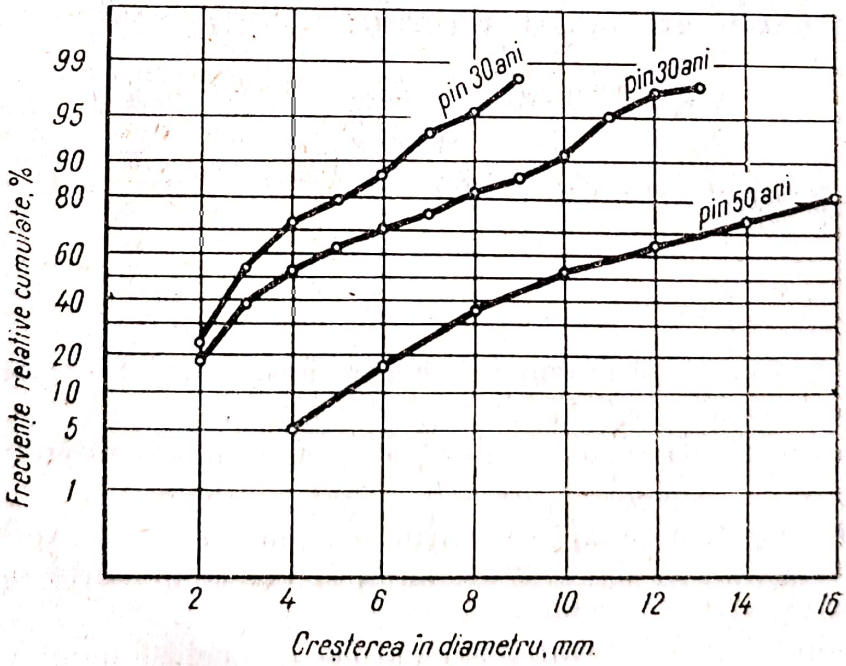


Fig. 3.10. Repartiția arborilor în raport cu creșterea în diametru, fără o grupare a arborilor pe categorii de diametre, pentru trei arborete echiene de pin (G i u r g i u, 1957)

Scanned with OKEN Scanner

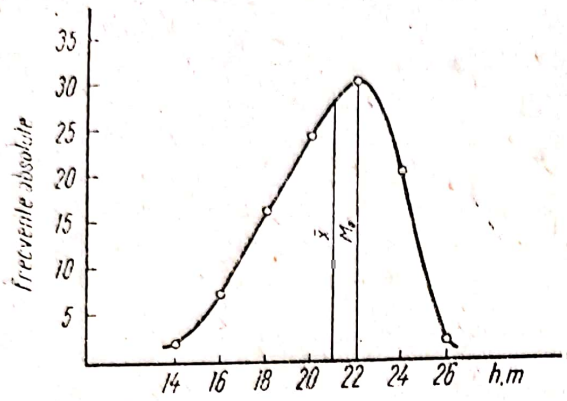


Fig. 3.11. Asimetrie de dreapta (Distribuția arborilor în raport cu înălțimea. Molid, 45 ani, Ocolul silvic Iacobeni, U.P. Orata, u.a. 27 a.).

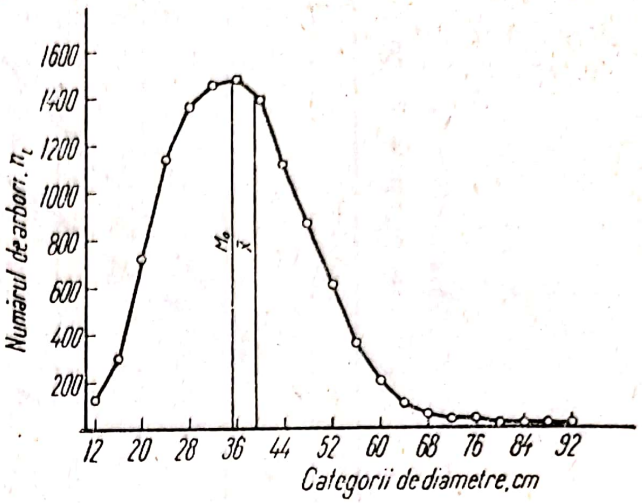


Fig. 3.12. Asimetrie de stînga (Funcția de frecevnță experimentală privind reparația arborilor pe categorii de diametre într-un arboret echien de brad, ocolul silvic Sinaia)

de faptul că arborii copleșiți sînt repede eliminați. Dezvoltarea liberă a arborilor groși determină o prelungire a ramurii drepte a curbei de distribuție.

În alte situații curbele pot fi simetrice dar să prezinte *excese*, respectiv în jurul mediei să rezulte frecvențe mai mari sau mai mici decît frecvența corespunzătoare distribuției normale. Excesul poate fi *pozitiv* (cînd frecvențele experimentale sînt mai mari decît cele date de funcția teoretică) sau *negativ* (cînd curba este mai teșită decît cea normală).

Măsura asimetriei este dată de raportul:

$$\alpha_3 = \frac{\mu_3}{(\sqrt{\mu_2})^3} \tag{3.23}$$

iar indicele excesului este dat de formula

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} - 3 \tag{3.24}$$

unde μ_2, μ_3 și μ_4 reprezintă momentele centrale de ordinul doi, trei și patru.

Pentru distribuția normală atît $\alpha_3 = 0$, cît și $\alpha_4 = 0$. Pentru $\alpha_3 < 0$ avem o asimetrie negativă de dreapta; $\alpha_3 > 0$ indică o asimetrie pozitivă (de stînga).

Dacă avem inegalitatea $\alpha_4 > 0$ curba este mai ascuțită spre vîrf decît curba normală și invers, dacă avem $\alpha_4 < 0$, curba prezintă o teșitură în comparație cu o curbă normală.

Pentru exemplul din tabelul 3.9 (echipat în același mod ca și tabelul 2.2) privind greutatea a 1 000 semințe de salcîm la 230 probe, obținem:

$$\alpha_3 = -0{,}71\,;\ \alpha_4 = 3{,}8$$

Scanned with OKEN Scanner

Calculul indicilor de asimetrie şi exces (x = **greutatea a** *1 000* **seminţe de salcîm, grame**)

x_0 =

Variabila x_i	Frecvenţele n_i	(1)	(2)	(3)	(4)
10,5	2	2	2	2	2
11,5	0	2	4	6	8
12,5	4	6	10	16	24
13,5	0	6	16	32	56
14,5	0	6	22	54	110
15,5	10	16	38	92	202
16,5	12	28	66	158	—
17,5	30	58	124	—	—
18,5	46	104	—	—	—
19,5	72	—	—	—	—
20,5	30	54	—	—	—
21,5	10	24	48	—	—
22,5	8	14	24	38	—
23,5	2	6	10	14	18
24,5	4	4	4	4	4
	230				
	a_i	102	86	56	22
	b_i	228	282	360	402
	S_i	330	368	416	424
	d_i	—126	—196	—304	—380

$$m_1 = \frac{-126}{230} = -0{,}55; \qquad m_2 = \frac{330 + 2 \cdot 368}{230} = 4{,}63; \qquad \bar{x} = 19{,}50 - 0{,}55 = 18{,}95 \text{ g};$$

$$h = 1: \qquad s = \sqrt{4{,}63 - 0{,}55^2} = \pm 2{,}08$$

$$m_3 = \frac{-126 + 6(-196) + 6(-304)}{230} = -13{,}6$$

$$m_4 = \frac{330 + 14 \times 368 + 36 \times 416 + 24 \times 424}{230} = 133{,}2$$

$$\mu_2 = 4{,}63 - 0{,}55^2 = 4{,}33$$

$$\mu_3 = -13{,}6 - 3\,(4{,}63)\,(-0{,}55) + 2\,(-0{,}55)^3 = -6{,}4$$

$$\mu_4 = 133{,}2 - 4\,(-6{,}4)\,(-0{,}55) + 6\,(4{,}63)\,(-0{,}55)^2 - 3\,(0{,}55)^4 = 127{,}2$$

$$\alpha_3 = \frac{-6{,}4}{\left(\sqrt{4{,}33}\right)^3} = -0{,}71; \; s_{\alpha_3} = \sqrt{\frac{6}{230}} = \pm 0{,}16; \; \frac{\alpha_3}{s_{\alpha_3}} = \frac{0{,}71}{0{,}16} = 4{,}4 > 2$$

$$\alpha_4 = \frac{127{,}2}{4{,}33^2} - 3 = 3{,}8; \; s_{\alpha 4} = \pm 0{,}32; \; \frac{\alpha_4}{s_{\alpha_4}} = \frac{3{,}8}{0{,}32} = 11{,}9 > 2$$

Scanned with OKEN Scanner

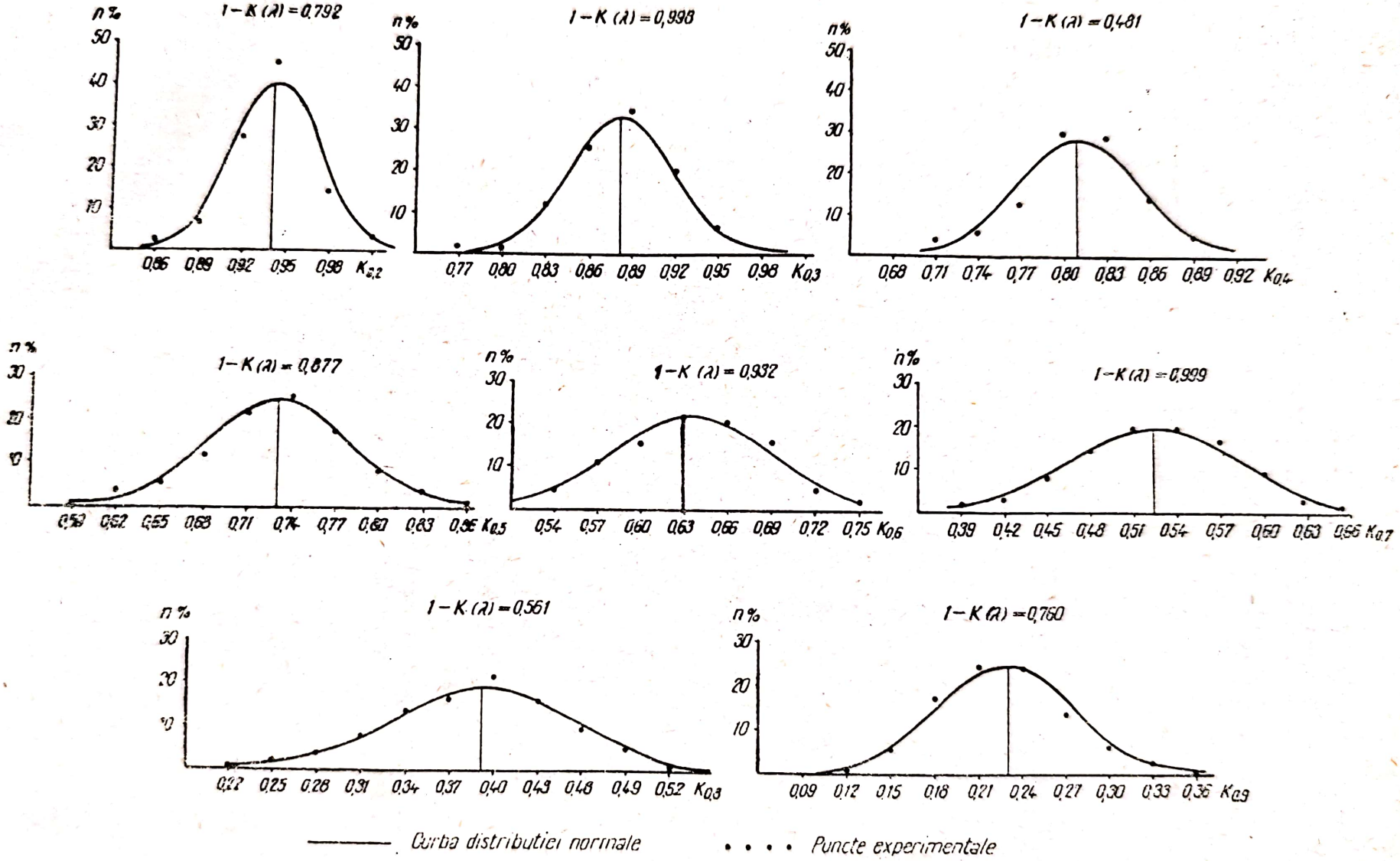


Fig. 3.13. Distribuții normale (Distribuția arborilor de molid din nordul țării în raport cu valorile indicilor de formă reali K_i; $i = 0,2; 0,3; ...0,9$)

Prezența asimetriei sau a excesului este dovedită dacă $\alpha_3 : s_{\alpha_3} > 2$, respectiv $\alpha_4 : s_{\alpha_4} > 2$ (pentru probabilitatea de 95%), unde $s_{\alpha_3} = \sqrt{\frac{6}{n}}$, iar $s_{\alpha_4} = 2\ s_{\alpha_3}$. Rezultă deci (tabelul 3.9) că distribuția luată în considerare prezintă atît asimetrie negativă (de dreapta) cît și exces pozitiv.

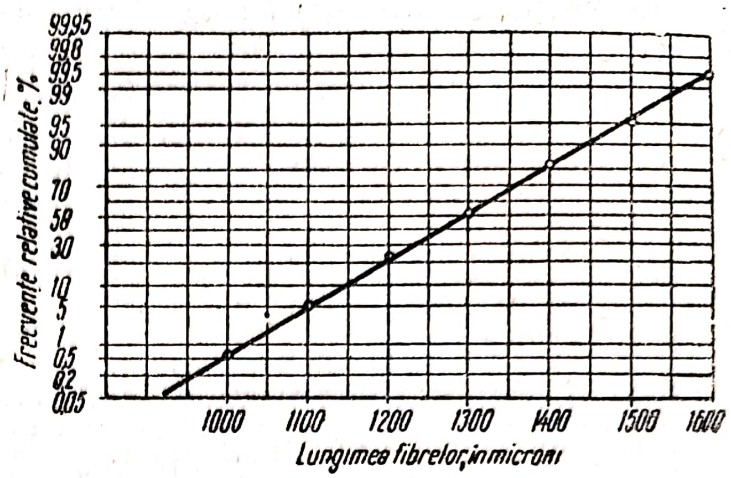


Fig. 3.14. Distribuție simetrică. Frecvențele relative cumulate în % ale lungimii fibrelor din lemn de nuc (prelucrare după Ghelmeziu și Istrate, 1960)

Prezența asimetriei sau a excesului nu trebuie înțeleasă în accepțiunea unui fenomen anormal. Din contră, în multe cazuri asimetria și excesul sînt specifice distribuției anumitor caracteristici, derivă din însăși natura fenomenului studiat și se pot explica din punct de vedere științific. Spre exemplu, asimetria este tipică pentru: distribuția diametrelor în arboretele echiene (frecvent, asimetrie de stînga); distribuția înălțimilor în arboretele echiene (asimetrie de dreapta); distribuția creșterilor în diametru (asimetrie de stînga); distribuția mărimii coroanelor (asimetrie de stînga) etc.

Mult mai simetrice și apropiate de curba normală sînt distribuțiile referitoare la însușirile morfologice ale plantelor, la caracteristicile biometrice ale semințelor forestiere, cele referitoare la indicii de formă ai arborilor (fig. 3.13), precum și distribuțiile privind proprietățile fizico-mecanice ale lemnului. În fig. 3.14 și 3.15 se prezintă pe hîrtie probabilistică frecvențele relative cumulate (în procente) privind lungimea fibrelor din lemnul de nuc și salcîm.

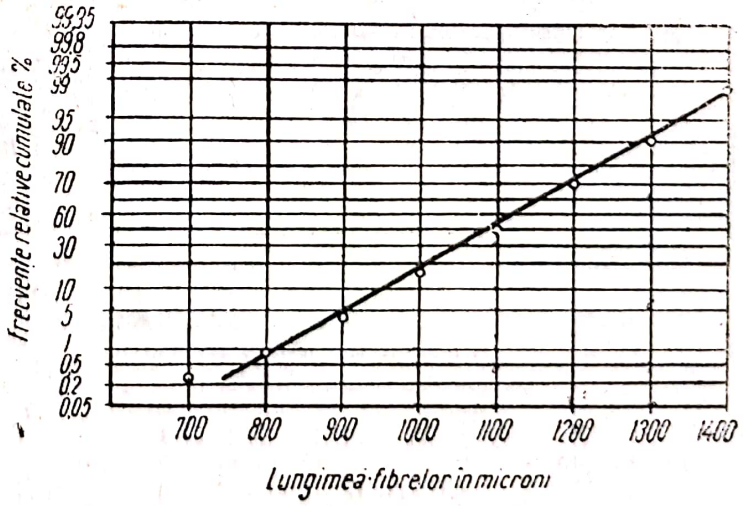


Fig. 3.15 Distribuție simetrică. Frecvențele relative cumulate (în %) ale lungimii fibrelor din lemnul de salcîm (prelucrare după Ghelmeziu și Istrate, 1960)

Scanned with OKEN Scanner

3.5. AJUSTAREA DISTRIBUȚIILOR ASIMETRICE

Pentru ajustarea distribuțiilor asimetrice se cunosc mai multe modalități. În cele ce urmează prezentăm una dintre acestea.

Ajustarea după formula CHURLIER, tipul A. Potrivit acestei formule, distribuția se calculează astfel:

$$\hat{n}_j = \psi(u)\frac{N}{s} = \left[f(u) - \frac{\alpha_3}{6}f^{III}(u) + \frac{\alpha_4}{24}f^{IV}(u)\right]\frac{N}{s} \qquad (3.25)$$

unde:

$f(u)$	— funcția de frecvență a distribuției normale (3.16)
$f^{III}(u)$ și $f^{IV}(u)$	— a treia și a patra derivată a funcției $f(u)$;
$\psi(u)$	— distribuția CHARJIER;
α_3	— indicele asimetriei;
α_4	— indicele excesului;
n_j	— frecvențe ajustate.

Funcția $f(u)$ reprezintă distribuția normală, al doilea membru reflectă influența asimetriei, iar al treilea membru exprimă influența excesului. Pentru $\alpha_3 = \alpha_4 = 0$

$$\psi(u) = f(u)$$

Prin urmare, formula distribuției normale reprezintă un caz particular al formulei CHARLIER, cînd $\alpha_3 = 0$ și $\alpha_4 = 0$.

Funcția $f(u)$, precum și funcțiile derivate $f^{III}(u)$ și $f^{IV}(u)$ se calculează în raport cu variabila normată u, după tabelul din anexa 2.

Pentru exemplificare prezentăm în tabelul 3.10 ajustarea unei serii statistice privind rezistența la compresiune paralel cu fibrele la lemnul de pin.

Prin metoda sumelor au rezultat (față de $x_0 = 415$):

$$m_1 = +0{,}041;\ m_2 = 3{,}157;\ m_3 = +1{,}490;\ m_4 = 30{,}271$$

$$\mu_2 = 3{,}155;\ \mu_3 = 1{,}102;\ \mu_4 = 30{,}059;$$

$$\mu_2' = 3{,}072;\quad \mu_4' = 28{,}510;\ \bar{x} = 416{,}64;$$

$$s = 1{,}752;\ \alpha_3 = +0{,}205;\ \alpha_4 = +0{,}021$$

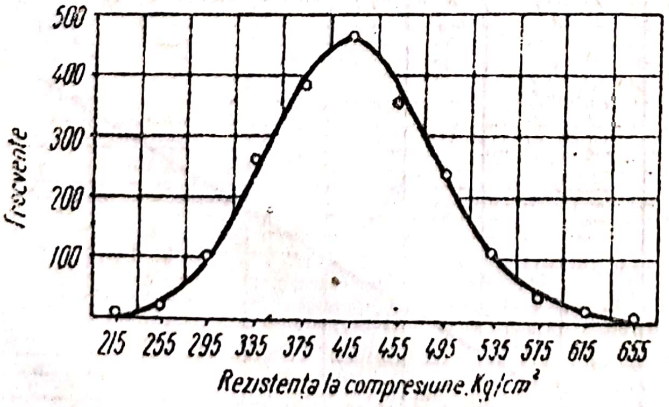


Fig. 3.16. Distribuția CHARJIER, tip A (Rezistența la compresiune pentru lemnul de pin silvestru)

Introducînd valorile lui α_3 și α_4 în formula (3.25) obținem:

$$\hat{n}_j = [f(u) - 0{,}034\ f^{III}(u) + 0{,}001\ f^{IV}(u)]\frac{2\,000}{1{,}752}$$

(Din anexa 2 ,pentru valori negative ale lui u în cazul funcției $f^{III}(u)$ se citesc cifrele cu semnul invers).

În fig. 3.16 se poate vedea cît de apropiate sînt punctele

Scanned with OKEN Scanner

Tabelul 3.10

Schema de calcul privind ajustarea distribuției empirice după formula CHARLIER, Tip A (Rezistența la compresiune paralel cu fibrele la lemnul de pin*)

x	$x'=\frac{x-x_0}{h}$	n_i	$x'-m_1$	$u=\frac{x'-m_1}{s}$	$f(u)$	$f^{III}(u)$	$f^{IV}(u)$	$-\frac{\alpha_3}{6}f^{III}(u)$	$\frac{\alpha_4}{24}f^{IV}(u)$	(5)+(8)+(9)	$n_j=\frac{N}{s}\times$ ×(10)
(0)	1	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
175	—6	—	—6,041	—3,446	0,00105	+0,03220	+0,7660	—0,00108	+0,00008	0,00005	0,06
215	—5	7	—5,041	—2,876	0,00638	+0,09672	+0,13899	—0,00329	+0,00014	0,00323	3,69
255	—4	22	—4,041	—2,305	0,02800	+0,14929	—0,01822	—0,00508	—0,00002	0,02290	26,13
295	—3	102	—3,041	—1,735	0,08857	+0,00146	—0,53132	—0,00005	—0,00053	0,08799	100,39
335	—2	260	—2,041	—1,164	0,20263	—0,38800	—0,66733	+0,01319	—0,00067	0,21515	245,46
375	—1	386	—1,041	—0,594	0,33382	—0,52582	+0,33691	+0,01788	+0,00034	0,35204	401,64
415	0	461	—0,041	—0,023	0,39883	—0,02751	+1,19518	+0,00094	+0,00120	0,40097	457,47
455	1	356	0,959	0,547	0,34350	+0,50746	+0,44458	—0,01725	+0,00044	0,32669	372,72
495	2	239	1,959	1,118	0,21355	+01,4782	—0,62720	—0,01420	—0,00063	0,19872	226,72
535	3	108	2,959	1,688	0,09598	+0,02441	—0,57374	—0,00083	—0,00057	0,09458	107,91
575	4	40	3,959	2,258	0,03117	—0,14771	—0,04976	+0,00502	—0,00005	0,03614	41,23
615	5	15	4,959	2,829	0,00729	—0,10325	+0,13884	+0,00351	+0,00014	0,01094	12,48
655	6	4	5,959	3,399	0,00123	—0,03594	+0,08304	+0,00122	+0,00008	0,00253	2,89
695	7	—	6,959	3,969	0,00016	—0,00790	+0,02436	+0,00027	+0,00002	0,00045	0,51
—	—	2 000	—	—	1,75216	—	—	—	—	1,75235	1999,30

* După M i t r o p o l s k i, 1961

— Pentru simplificarea calculelor se folosesc valorile reduse $x'=\frac{x-x_0}{h}$

Scanned with OKEN Scanner

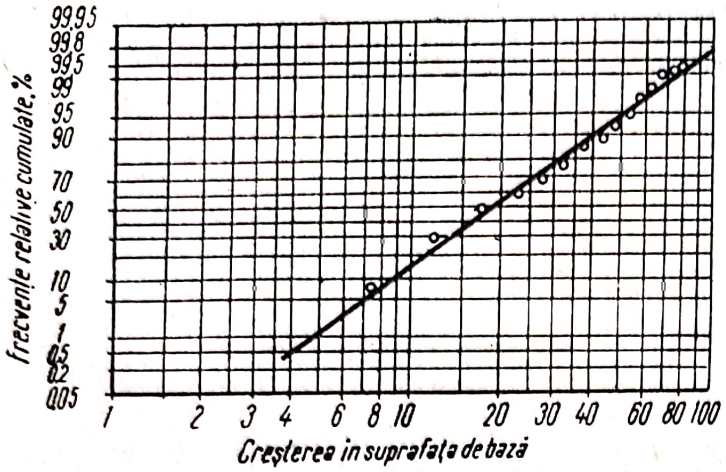


Fig. 3.17. Hîrtie probabilistică pentru distribuția normal-logaritmică (Repartiția arborilor în raport cu creșterea în suprafață de bază la un arboret echien de molid — Giurgiu, 1957)

experimentale de curba teoretică, ceea ce dovedește că fenomenul studiat urmează legea distribuției CHARLIER.

Formula CHARLIER poate fi cu succes aplicată în biometria forestieră pentru ajustarea multor distribuții. Rezultate deosebit de corecte se obțin în ceea ce privește distribuția arborilor pe categorii de diametre, precum și în multe alte probleme din domeniul biometriei forestiere.

Pentru ajustarea unor distribuții experimentale pot fi folosite și alte funcții, cum sînt cele propuse de PEARSON, distribuția normal-logaritmică (fig. 3.17), distribuția MAXWELL, distribuția binomial-negativă, distribuția exponențială etc. (P r o d a n, 1961; H a l d, 1952; D l i n, 1958; M i t r o p o l s k i, 1961; ș.a.).

3.6. ERORILE PRINCIPALILOR PARAMETRI AI DISTRIBUȚIILOR STATISTICE UNIDIMENSIONALE

De cele mai multe ori, principalii parametri ai distribuțiilor statistice se determină pe baza unui număr limitat de valori observate, din motive practice fiind imposibil a se cerceta toate unitățile colectivității luate în studiu. De pildă, germinația tehnică a unui lot de semințe se stabilește dintr-un număr limitat de semințe luate din lotul respectiv; de aceea media calculată va fi afectată de unele erori care provin în principal din cauză că nu s-au luat în considerare toate semințele lotului dat. De aceleași erori sînt afectate și valorile altor parametri ai distribuțiilor statistice (abaterea medie pătratică, coeficientul de variație, indicele asimetriei etc.).

În statistica matematică se cunosc următoarele formule pentru determinarea acestor erori:

— eroarea standard a mediei aritmetice (eroarea medie a mediei aritmetice)

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{3.26}$$

sau, în procente:

$$s_{\bar{x}\%} = \frac{s\%}{\sqrt{n}} \tag{3.27}$$

— eroarea standard a abaterii medii pătratice

$$s_s = \frac{s}{\sqrt{2n}} \tag{3.28}$$

— eroarea standard a coeficientului de variație

$$s_{s_{\%}} = \frac{s_{\%}\sqrt{0{,}5 + \left(\frac{s_{\%}}{100}\right)^2}}{\sqrt{n}} \tag{3.29}$$

— eroarea standard a indicelui de asimetrie

$$s_{\alpha_3} = \sqrt{\frac{6}{n}} \tag{3.30}$$

— eroarea standard a indicelui de exces

$$s_{\alpha_4} = 2\, s_{\alpha_3} \tag{3.31}$$

— eroarea standard a proporției

$$s_p = \sqrt{\frac{p(1-p)}{n}} \tag{3.32}$$

Aceste erori arată limitele în care cu probabilitatea de 0,683 poate fi întîlnită valoarea parametrului respectiv.

După cum rezultă din formulele prezentate, erorile sînt invers proporționale cu rădăcina pătrată a numărului de valori observate. Pentru un număr mare de valori observate erorile vor fi foarte mici, iar indicatorii statistici determinați pe baza colectivității de selecție se apropie de indicatorii colectivității generale (vezi capitolul 8).

Pentru o distribuție statistică cu 230 valori observate, reprezentînd variația greutății semințelor de salcîm, au rezultat următorii indicatori statistici (tabelul 3.9):

$$\bar{x} = 18{,}95; \quad s = \pm 2{,}08; \quad s_{\%} = \pm 11\%; \quad \alpha_3 = -0{,}71; \quad \alpha_4 = 3{,}8.$$

Erorile acestor indicatori vor fi:

$$s_{\bar{x}} = \frac{2{,}08}{\sqrt{230}} = 0{,}138; \quad s_{\bar{x}\%} = \frac{11}{\sqrt{230}} = 0{,}73\%$$

$$s_s = \frac{2{,}08}{\sqrt{2\times 230}} = 0{,}097; \quad s_{s\%} = \frac{11\cdot\sqrt{0{,}5+\left(\frac{11}{100}\right)^2}}{\sqrt{230}} = 0{,}52$$

$$s_{\alpha_3} = \sqrt{\frac{6}{230}} = \pm 0{,}16; \quad s_{\alpha_4} = 2\times 0{,}16 = \pm 0{,}32$$

Potrivit celor de mai sus se pot stabili limite pentru principalii indicatori statistici ai distribuției studiate. Astfel, în ceea ce privește media aritmetică rezultă că în 683 din cazuri la mie, media reală se situează în intervalul 18,81 și 19,09, iar în 954 din cazuri în intervalul 18,68 și 19,22.

Scanned with OKEN Scanner

IV

EXEMPLE DIN SILVICULTURĂ PRIVIND VERIFICAREA IPOTEZELOR STATISTICE

4.1. CONSIDERAȚII GENERALE

În economia forestieră multe cercetări experimentale se efectuează în scopul estimării parametrilor distribuțiilor statistice. De pildă, interesează fondul de producție la hectar, creșterea medie în înălțime a puieților plantați, procentul mediu de reușită al culturilor forestiere etc. În numeroase cazuri însă ,cercetările experimentale folosesc estimațiile parametrilor statistici în scopul verificării anumitor ipoteze. Așa, de exemplu, vrem să cunoaștem: dacă o nouă clonă de plopi euramericani este mai bună decît clona introdusă anterior în producție; dacă o nouă metodă de prelucrare a solului este mai eficientă decît metoda actuală; dacă aplicarea îngrășămintelor chimice are sau nu o influență asupra productivității arboretelor etc.

Se știe că în acest scop cercetătorul, în majoritatea cazurilor, se mulțumește cu o mică fracțiune a colectivității pe care o are de studiat. Este constrîns să procedeze astfel fiindcă întreaga colectivitate este practic inaccesibilă. De pildă, dacă am vrea să studiem proprietățile fizice și mecanice ale lemnului de molid, chiar dacă ne vom limita la arboretele exploatabile, vom avea de parcurs o colectivitate care ar depăși orice posibilități practice de realizare. Uneori, chiar dacă am putea cuprinde întreaga colectivitate, nu putem studia nici măcar o parte relativ mai mică a acesteia, fiindcă particularitățile anumitor cercetări în cadrul cărora se distrug materialele respective împiedică realizarea unui număr mare de valori observate. Acesta este cazul întocmirii tabelelor de cubaj, al studiului germinației semințelor forestiere etc., cînd luarea în considerare a întregii populații ar conduce la însemnate pagube materiale. De aceea, se renunță la observarea întregii populații și se trece la efectuarea de măsurători pe eșantioane. Apoi, de la parametrii eșantioanelor se trece la caracterizarea întregii colectivități. Cu cît volumul eșantionului este mai mare, cu atît eroarea de eșantionaj este mai mică. Desigur, la probele de volum redus, concluziile sînt mai puțin sigure, dar și în acest caz interpretarea rezultatelor este posibilă și utilă, arătînd însă gradul de nesiguranță al acestor concluzii.

Scanned with OKEN Scanner

Mediile tuturor eșantioanelor cercetate se grupează în jurul unei valori centrale a. Mediile de eșantionaj reprezintă doar estimații ale valorii centrale a. De aceea, niciodată nu se afirmă că media unei colectivități este a, ci numai că această medie este mai mult sau mai puțin apropiată de a, variind în cadrul unui *interval de încredere*. Nu este corect să se considere că estimația obținută este însăși valoarea parametrului căutat, așa cum mulți cercetători sînt tentați să interpreteze. Estimația însă dă posibilitatea să se determine un interval care acoperă cu o anumită probabilitate cunoscută valoarea parametrului estimat. Se spune că media (sau alt parametru al distribuțiilor statistice) este mai mare decît a_i, dar mai mică decît a_s, fără a putea preciza valoarea exactă a acestui parametru. Fiecare interval de încredere este legat de o anumită *siguranță statistică* (siguranță probabilistică) și de un anumit *nivel de semnificație q* (probabilitate de transgresiune, prag de semnificație). Atunci cînd se afirmă că „media este cuprinsă între a_i și a_s cu o probabilitate de 99%", înțelegem că studiînd un număr mare de eșantioane din aceeași populație în 99% din cazuri media nu va depăși limitele menționate. Aceste limite pot fi totuși depășite în 1% din cazuri. Deci, riscul de a greși este de 1%. S-a convenit să se spună că nivelul de semnificație este de 1% sau de 0,01. Pentru un nivel de semnificație de 0,05 riscul de a greși este mai mare, respectiv siguranța statistică este mai mică (0,95). O siguranță statistică mai mică de 95% (nivel de semnificație mai mare de 0,05) este considerată ca fiind necorespunzătoare.

La interpretarea rezultatelor se ridică probleme de comparație între doi parametri (două medii, două abateri standard etc.), între două distribuții, între un parametru și valoarea lui teoretică etc. La rezolvarea acestor probleme se pune așa-numita *ipoteză nulă* (H_o), potrivit căreia eșantioanele comparate sînt extrase din aceeași colectivitate generală și ca atare diferența dintre parametri se datorește întîmplării, în realitate diferența fiind *nulă*. Plecînd de la această ipoteză se calculează probabilitatea de a se extrage din aceeași colectivitate două selecții (eșantioane) între ai căror parametri să existe o diferență egală sau mai mare decît diferența stabilită experimental.

În cazul în care această probabilitate este mare, se acceptă ipoteza nulă, admițîndu-se că diferențele constatate experimental se datoresc întîmplării și ca atare acestea *nu sînt semnificative*. Se spune că cele două eșantioane sînt extrase din aceeași colectivitate. În caz contrar, diferențele sînt considerate *semnificative* și prin aceasta se respinge ipoteza nulă. Se afirmă că este puțin probabil ca cele două eșantioane să fie extrase din aceeași colectivitate. Respingînd ipoteza nulă, cercetătorul poate greși respingînd o ipoteză adevărată (se comite o eroare de genul I). Admițînd că o diferență este întîmplătoare, cînd în realitate ea nu este, se comite o eroare de genul II.

Trebuie deci respinse numai ipotezele care se pot realiza cu o probabilitate foarte mică. S-a convenit ca să se respingă ipotezele care se pot realiza cu o probabilitate mai mică de 0,05 (5%), iar diferența se consideră semnificativă. Unii cercetători consideră semnificative numai diferențele corespunzătoare unei probabilități mai mici de 0,01 sau 0,001. Alegerea acestui nivel de semnificație depinde de importanța problemei, de exigența cercetătorului, precum și de posibilitățile practice de realizare a unui asemenea nivel.

Scanned with OKEN Scanner

Legătura dintre intervalul de încredere, siguranța statistică (siguranța probabilistică) și nivelul de semnificație se poate urmări în fig. 4.1.

Cunoscînd că abaterea standard a distribuției de medii (eroarea standard a mediei aritmetice) este de $\frac{\sigma}{\sqrt{n}}$ (unde σ este abaterea standard a populației) și însemnînd cu $\bar{x}$ — media eșantionului și cu a — media populației, putem stabili intervalele de încredere pentru diferite nivele de semnificație. Astfel:

— pentru nivelul de semnificație 0,05

$$\bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}} < a < \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}};$$

— pentru nivelul de semnificație 0,01

$$\bar{x} - 2{,}58\frac{\sigma}{\sqrt{n}} < a < \bar{x} + 2{,}58\frac{\sigma}{\sqrt{n}};$$

— pentru nivelul de semnificație 0,001

$$\bar{x} - 3{,}29\frac{\sigma}{\sqrt{n}} < a < \bar{x} + 3{,}29\frac{\sigma}{\sqrt{n}}.$$

dau, generalizînd:

$$\bar{x} - u\frac{\sigma}{\sqrt{n}} < a < \bar{x} + u\frac{\sigma}{\sqrt{n}}$$

unde u se stabilește în raport cu nivelul de semnificație dorit după tabelul din anexa 3. Astfel, pentru nivelul de semnificație de 0,05 avem $u = 1{,}96$.

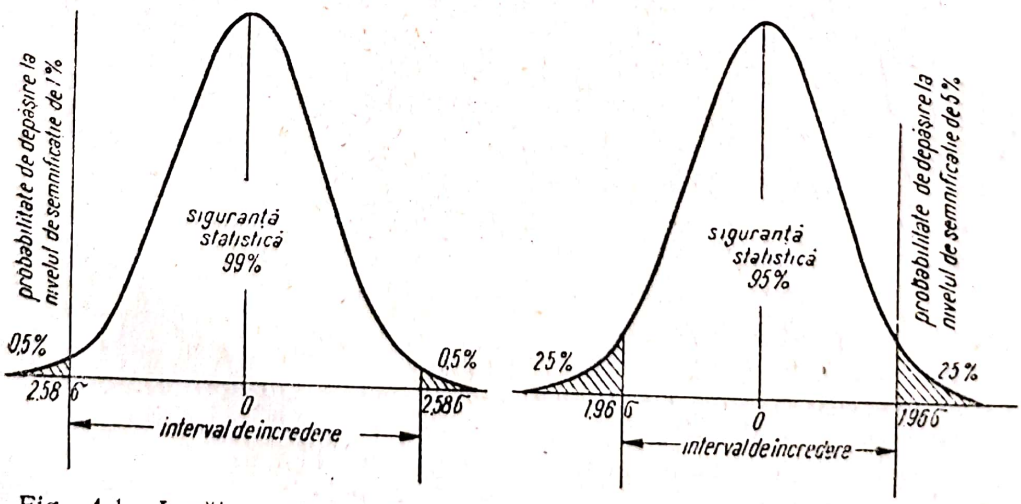


Fig. 4 1. Legătura dintre intervalul de încredere, siguranța statistică și nivelul de semnificație

Scanned with OKEN Scanner

Cazul colectivităţilor de volum mare

Stabilirea semnificaţiei diferenţelor între două medii este de cea mai mare importanţă pentru cercetările din domeniul economiei forestiere, unde întîlnim o mare variabilitate a caracteristicilor studiate.

Se ştie că la repetarea experienţelor prin sondaje, mediile obţinute diferă între ele. Distribuţia acestor valori medii este foarte apropiată de distribuţia normală. Gradul de împrăştiere al mediilor individuale în jurul mediei colectivităţii generale este dat de abaterea medie pătratică a mediei aritmetice, denumită eroarea standard a mediei aritmetice. Aceasta se calculează după formula:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{3.25}$$

Media de selecţie are deci o „instabilitate de eşantionaj". Este firesc ca această instabilitate să se întîlnească şi la diferenţele dintre două medii Gradul de instabilitate al acestor diferenţe este dat de *eroarea diferenţelor* care se calculează astfel:

$$s_d = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}} \tag{4.1}$$

sau:

$$s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{4.2}$$

unde n_1 şi n_2 sînt mai mari de 30.

Prin raportarea diferenţei dintre două medii $(\bar{x}_1 - \bar{x}_2)$ la eroarea diferenţei s_d se obţin valorile u, ca şi în cazul distribuţiei valorilor individuale.

Prin urmare

$$u = \frac{\bar{x}_1 - \bar{x}_2}{s_d} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{4.3}$$

Diferenţa dintre $\bar{x}_1$ şi $\bar{x}_2$ este cu atît mai semnificativă cu cît u este mai mare, respectiv cu cît s_d este mai mic şi cu cît $(\bar{x}_1 - \bar{x}_2)$ este mai mare. Pentru valori ale lui $u > 2$ este foarte mică probabilitatea ca diferenţa $(\bar{x}_1 - \bar{x}_2)$ să fie datorită instabilităţii de eşantionaj şi ca atare trebuie atribuită unor cauze reale, respectiv efectului anumitor factori sau tratamente aplicate.

În cazul cînd este necesar să se compare o medie cu o valoare cunoscută a, în ipoteza distribuţiei normale, vom aplica formula:

$$u = \frac{\bar{x} - a}{s}\sqrt{n} \tag{4.4}$$

unde x şi s sînt parametrii distribuţiei respective, avînd n valori observate $(n > 30)$, iar u este testul (criteriul) pentru stabilirea gradului de semnificaţie.

Scanned with OKEN Scanner

Cazul colectivităţilor de volum redus. Sînt situaţii cînd este necesară compararea a două medii, reprezentînd două colectivităţi de selecţie de volum redus ($n < 30$). Asemenea probleme se rezolvă în mod diferenţiat, după cum dispersiile colectivităţilor generale din care s-au extras selecţiile sînt sau nu egale. Egalitatea dispersiilor σ_1^2 şi σ_2^2 se poate verifica după indicaţiile date la § 4.3.

In cazul cînd σ_1^2 şi σ_2^2 sînt considerate egale, verificarea ipotezei egalităţii celor două medii se realizează cu ajutorul formulei:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{4.5}$$

unde:

$$s^2 = \frac{f_1 s_1^2 + f_2 s_2^2}{f_1 + f_2} \tag{4.6 a}$$

unde $f_1 = n_1 - 1$; $f_2 = n_2 - 1$; $f_1 + f_2 = n_1 + n_2 - 2$.

Cînd $n_1 = n_2$

$$s^2 = \frac{s_1^2 + s_2^2}{2} \tag{4.6 b}$$

Valoarea t calculată se compară cu valoarea t tabelară (anexa 4) la $f = n_1 + n_2 - 2$ *grade de liberate*, în raport cu nivelul de semnificaţie luat în considerare.

Pentru *situaţiile cînd σ_1^2 şi σ_2^2 nu sînt considerate egale*, se aplică formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{4.7}$$

unde t urmează distribuţia STUDENT (anexa 4), cu numărul gradelor de libertate determinat după următoarea formulă:

$$f = \frac{1}{\frac{C^2}{n_1 - 1} + \frac{(1 - C)^2}{n_2 - 1}} \tag{4.8}$$

unde

$$C = \frac{\frac{s_1^2}{n_1}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \tag{4.9}$$

Dacă $n_1 = n_2$ avem:

$$f = \frac{n - 1}{C^2 + (1 - C)^2} \tag{4.10}$$

Scanned with OKEN Scanner

unde

$$C = \frac{s_1^2}{s_1^2 + s_2^2} \qquad (4.11)$$

Aşadar, înainte de a trece la verificarea egalităţii a două medii pentru selecţii de volum mic se impune o verificare a ipotezei $\sigma_1^2 = \sigma_2^2$, adică se verifică dacă dispersiile celor două colectivităţi generale din care au fost extrase selecţiile sînt sau nu egale. Această verificare se realizează prin intermediul criteriilor F (§ 4.3).

Cînd se compară o medie de selecţie cu o valoare cunoscută (constantă) testul t se calculează după formula (4.4), iar numărul gradelor de libertate va fi $n - 1$.

*

Diferenţele se consideră:

— *semnificative*, atunci cînd corespund unei probabilităţi de transgresiune (de depăşire) q de cel mult 5% (0,05);

— *distinct semnificative*, atunci cînd corespund unei probabilităţi q de cel mult 1%;

— *foarte semnificative*, atunci cînd corespund unei probabilităţi q mai mici de 0,1%.

În cazul unor selecţii de volum mare, aceste probabilităţi corespund următoarelor valori u:

Diferenţe	Nivelul de semnificaţie. (Probabilitatea de transgresiune) %	u
Semnificative	$\leqslant 5$	$\geqslant 1{,}96$
Distinct semnificative	$\leqslant 1$	$\geqslant 2{,}58$
Foarte semnificative	$\leqslant 0{,}1$	$\geqslant 3{,}29$

Pentru alte nivele de semnificaţie valorile u se găsesc în anexa 3.

Atunci cînd numărul valorilor observate este redus, se folosesc valorile t corespunzătoare nivelurilor de semnificaţie de 5%, 1% şi 0,1%. Aceste valori se dau în anexa 4 (valorile distribuţiei STUDENT). Intrarea în tabel este dată de numărul *gradelor de libertate*.

Cu valorile din anexa 4 se compară valorile t găsite experimental. Cînd valorile t experimentale sînt mai mari decît valorile t trecute în acest tabel pentru $q = 5\%$, 1% sau 0,1%, diferenţele dintre medii vor fi considerate, respectiv semnificative, distinct semnificative şi foarte semnificative.

Pentru exprimarea semnificaţiei statistice a diferenţelor, uneori se folosesc următoarele simboluri:

	Diferenţe	
	Pozitive	*Negative*
Semnificative	*	0
Distinct semnificative	* *	0 0
Foarte semnificative	* * *	0 0 0

Exemplul 1. Studiind valorile indicilor calitativi ai semințelor de brad (Enescu — Costea, 1962) *, s-a urmărit să se stabilească dacă condițiile climatice generale deosebite din Oltenia și Banat determină sau nu o variație a calității. În acest scop, s-au luat în studiu 167 probe provenite din Oltenia, din sectorul climatic IV-F și 210 probe provenite din Banat, din sectorul climativ IV-E.

În final s-au obținut rezultatele din tabelul 4.1.

Tabelul 4.1

Indicele calitativ	Oltenia		Banat	
	$\bar{x}$	$s_{\bar{x}}$	$\bar{x}$	$s_{\bar{x}}$
Potența germinativă, %	37,55	0,76	50,82	1,06
Semințe pline, %	55,78	0,79	66,52	0,79
Greutatea a 1 000 de semințe, %	60,57	0,75	69,64	0,75

Aplicînd formula (4.3) obținem:

— pentru potența germinativă:

$$u = \frac{50,82 - 37,55}{\sqrt{1,06^2 + 0,76^2}} = 10,2$$

— pentru procentul semințelor pline:

$$u = \frac{66,52 - 55,78}{\sqrt{0,79^2 + 0,79^2}} = 9,6$$

— pentru greutatea a 1 000 de semințe:

$$u = \frac{69,64 - 60,57}{\sqrt{0,75^2 + 0,75^2}} = 8,5$$

Pentru toți acești indici au rezultat valori u mai mari decît 3,29, ceea ce arată că diferențele sînt foarte semnificative. Această concluzie, fundamentată temeinic prin calcule statistico-matematice, a scos în evidență necesitatea stabilirii unor indici regionali.

Exemplul 2. La cîte 12 loturi de manele s-a calculat volumul prin două metode de cubaj: 1) după tabela volumului cilindrului în funcție de diametrul la mijlocul lungimii și 2) după tabela de cubaj specială, în funcție de diametrul măsurat la capătul subțire al pieselor. Față de volumul exact, s-au obținut rezultatele (în %) menționate în tabelul 4.2.

* Revista Pădurilor, nr. 12, 1962

Se pune în mod firesc întrebarea dacă cele două metode dau rezultate identice și care din ele este mai bună pentru aplicații practice?

Pentru a stabili dacă rezultatele unei metode diferă datorită unor erori sistematice și nu întîmplător, urmează să aplicăm formula (4.5), corespunzătoare colectivităților de volum mic (σ_1^2 și σ_2^2 sînt considerate egale. vezi exemplul 2, § 4.3).

Calculăm în primul rînd mărimea s^2, după formula (4.6 b), deoarece $n_1 = n_2$;

$$s^2 = \frac{s_1^2 + s_2^2}{2} = \frac{11{,}3 + 23{,}7}{2} = 17{,}5;$$

$$s = 4{,}2$$

unde:

$$s_1^2 = \frac{\sum x_1^2 - \frac{\Sigma(x_1)^2}{n_1}}{n_1 - 1} = 11{,}3$$

$$s_2^2 = \frac{\sum x_2^2 - \frac{\Sigma(x_2)^2}{n_2}}{n_2 - 1} = 23{,}7$$

Tabelul 4.2

metoda 1 x_{1i}	metoda 2 x_{2i}
96	102
107	101
97	105
98	97
96	103
99	103
101	95
98	101
95	104
104	99
98	108
98	90
$\bar{x}_1 = 99{,}0$	$\bar{x}_2 = 100{,}7$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{100{,}7 - 99{,}0}{4{,}2\sqrt{\frac{1}{12} + \frac{1}{12}}} = \frac{100{,}7 - 99{,}0}{4{,}2\sqrt{2}}\sqrt{12} = 0{,}99$$

Din anexa 4 rezultă că pentru a avea o probabilitate de transgresiune de 5%, la compararea a două metode cu 12 repetiții ($f = 12 + 12 - 2 = 22$ grade de libertate), diferența trebuie să fie de 2,07 ori mai mare decît eroarea diferenței (t tabelar = 2,07).

În exemplul luat, valoarea experimentală t este mai mică decît 2,07 și ca atare diferența dintre cele două medii trebuie considerată ca nesemnificativă. De aici concluzia: pe ansamblu cele două metode de cubaj sînt identice din punctul de vedere al preciziei.

Tabelul 4.3

pH determinat prin			
procedeul I	procedeul II	x_{1i}^2	x_{2i}^2
x_{1i}	x_{2i}		
4,5	5,5	20,2	30,2
6,5	5,5	40,2	30,2
5,0	5,5	25,0	30,2
5,5	5,5	30,2	30,2
6,3	6,5	39,7	42,2
6,0	5,5	36,0	30,2
6,0	5,5	36,0	30,2
6,3	6,5	39,7	42,2
5,5	5,5	30,2	30,2
5,0	5,5	25,0	30,2
6,5	5,5	42,2	30,2
4,6	5,5	20,2	30,2
67,6	68,0	386,6	386,4

Exemplul 3. Aciditatea solului (pH) se poate determina prin două procedee. Se pune întrebarea dacă aceste procedee dau rezultate diferite. În acest scop s-au luat cîte 12 probe de sol, obținînd datele din tabelul 4.3.

Scanned with OKEN Scanner

În baza datelor obținute rezultă:

$$\bar{x}_1 = 5{,}64 \qquad \bar{x}_2 = 5{,}67$$

$$s_1^2 = \frac{386{,}6 - \frac{(67{,}6)^2}{12}}{12-1} = \frac{5{,}8}{11} = 0{,}53$$

$$s_2^2 = \frac{386{,}4 - \frac{(68{,}0)^2}{12}}{12-1} = \frac{1{,}1}{11} = 0{,}10$$

Prin procedeul descris la § 4.3 s-a dovedit că $\sigma_1^2 \neq \sigma_2^2$

Prin urmare vom aplica formula (4.7). În acest scop, efectuăm următoarele calcule prealabile:

$$C = \frac{s_1^2}{s_1^2 + s_2^2} = \frac{0{,}53}{0{,}53 + 0{,}10} = \frac{0{,}53}{0{,}63} = 0{,}84$$

$$f = \frac{n-1}{C^2 + (1-C)^2} = \frac{12-1}{0{,}84^2 + (1-0{,}84)^2} = \frac{11}{0{,}74} = 15$$

Așa încît

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{5{,}64 - 5{,}67}{\sqrt{\frac{0{,}53}{12} + \frac{0{,}10}{12}}} = 0{,}13$$

Valoarea tabelară a testului t (f=15; q=0,05; t=2,45) este mult superioară valorii calculate a lui t, ceea ce dovedește că nu avem dovezi pentru respingerea ipotezei nule, respectiv diferența dintre cele două medii nu este semnificativă. Diferențele semnalate dintre cele două medii se datoresc factorilor întîmplători. Dispersiile sînt însă diferite (§ 4.3, exemplul 3).

Comportarea mediilor prin metoda cuplurilor. Frecvent, în cercetările din domeniul economiei forestiere și în activitatea practică se pune problema comparării a două metode sub raportul preciziei sau al eficacității. În acest scop, așa cum s-a arătat, se folosește testul t, ca și în cazul comparării mediilor, aplicînd formulele prezentate. Sînt însă situații cînd unitățile nu prezintă omogenitate sub raportul caracteristicii cercetate. Apoi sînt rare cazurile cînd se cunoaște în prealabil dacă este asigurată sau nu omogenitatea unităților luate în considerare. De aceea, este mai rațional să se extragă *perechi de unități* (cupluri) sau să se efectueze determinări folosind *una și aceeași probă* (R a n c u și T ö v i s s i, 1963). În acest mod se creează posibilitatea studierii efectelor celor două metode de cercetare în condiții cît mai diferite, prin eliminarea influenței unor

* Se folosește formula (2.39), § 2.5

factori secundari, permiţînd formularea unor concluzii cu o valabilitate mai generală. De data aceasta testul t se calculează astfel:

$$t = \frac{\bar{d}}{s_d}\sqrt{n} \tag{4.12}$$

unde

$$\bar{d} = \frac{\Sigma(x_{i1} - x_{i2})}{n} = \frac{\Sigma d_i}{n} \tag{4.13}$$

$$s_d^2 = \frac{\Sigma(d_i - \bar{d})^2}{n-1} = \frac{\sum d_i^2 - \frac{(\Sigma d_i)^2}{n}}{n-1} \tag{4.14}$$

Numărul gradelor de libertate este dat de relaţia: $f = n - 1$.

E x e m p l u. La 10 arborete de molid s-a determinat volumul pe picior prin următoarele două metode: o dată prin metoda tabelelor generale de cubaj, a doua oară prin metoda seriilor de volume* (tabelul 4.4).

Tabelul 4.4

Rezultatele determinării volumelor la 10 arborete de molid prin două metode de cubaj (m³/ha)

Arboretul	Metoda tabelelor generale de cubaj x_{i1}	Metoda seriilor de volume x_{i2}	d_i	d_i^2
1	101	103	2	4
2	155	150	—5	25
3	180	188	8	64
4	220	210	—10	100
5	270	272	2	4
6	310	315	5	25
7	390	380	—10	100
8	450	455	5	25
9	502	508	6	36
10	600	602	2	4
Σ	—	—	+5	387

Pentru a aprecia dacă diferenţa dintre rezultatele obţinute prin cele două metode este sau nu semnificativă, aplicăm formula (4.12)

Din tabelul 4.4 obţinem:

$$\bar{d} = \frac{5}{10} = 0{,}50$$

$$s_d^2 = \frac{387 - \frac{(5)^2}{10}}{9} = 42{,}9$$

$$t = \frac{0{,}5}{\sqrt{42{,}8}}\sqrt{10} = 0{,}2; \qquad f = n - 1 = 10 - 1 = 9$$

* G i u r g i u V,. Serii de volume pentru arboretele echiene, Revista pădurilor, nr. 11. 1964

Scanned with OKEN Scanner

Pentru $f=9$ și $q=0,05$, găsim în tabelul din anexa 4, $t=2,26$, care este mai mare decît t experimental. Conchidem deci că nu avem niciun temei pentru respingerea ipotezei nule. Metoda seriilor de volume dă practic aceleași rezultate ca și metoda tabelelor generale de cubaj; abaterile care se observă se datoresc factorilor întîmplători.

Compararea a două frecvențe (procente). Uneori se pune problema comparării a două procente reprezentînd frecvențele unui anumit caracter în două eșantioane. Să presupunem că în primul eșantion cu un efectiv n_1 s-a găsit caracterul respectiv în a_1 cazuri, iar în al doilea eșantion, format din n_2 unități, s-a găsit în a_2 cazuri. De aici rezultă frecvențele:

$$p_1 = \frac{a_1}{n_1} \quad \text{și} \quad p_2 = \frac{a_2}{n_2} \tag{4.15}$$

Semnificația diferenței dintre aceste două frecvențe se stabilește ușor, cunoscîndu-se că diferența $p_1 - p_2$ este distribuită normal în jurul unei medii nule. În acest scop se folosește criteriul u calculat după formula:

$$u = (p_1 - p_2)\sqrt{\frac{n_1 n_2}{pq(n_1 + n_2)}} \tag{4.16}$$

unde:

$$p = \frac{a_1 + a_2}{n_1 + n_2} \tag{4.17}$$

iar

$$q = 1 - p$$

Diferența este semnificativă (se respinge ipoteza nulă) atunci cînd u experimental este mai mare decît u tabelar corespunzător nivelului de semnificație ales. Pentru nivelul de semnificație de 1%, u experimental trebuie să fie mai mare de 2,58, iar pentru nivelul de semnificație de 5% avem nevoie de un $u > 1,96$ (anexa 3).

Exemplu. Din 500 insecte tratate cu insecticidul A au rezistat 20, iar din 400 insecte tratate cu insecticidul B au rămas 30. Se pune întrebarea dacă diferența dintre efectele celor două insecticide este statistic asigurată.

Trecînd la calcule avem:

$$n_1 = 500; \quad n_2 = 400; \quad a_1 = 20; \quad a_2 = 30$$

$$p = \frac{20 + 30}{500 + 400} = 0,055; \qquad q = 0,945$$

$$p_1 = \frac{20}{500} = 0,040; \qquad p_2 = \frac{30}{400} = 0,075$$

$$u = (0,075 - 0,040)\sqrt{\frac{500 \times 400}{0,040 \times 0,075(500+400)}} = 3,0$$

Pentru nivelul de semnificație de 1% (u tabelar de 2,58) diferența dintre efectele celor două insecticide este statistic asigurată. Potrivit celor arătate, diferența este distinct semnificativă.

Scanned with OKEN Scanner

La compararea unei frecvenţe a/n cu un parametru p cunoscut, vom folosi formula:

$$u = \frac{(a/n) - p}{\sqrt{\frac{p(1-p)}{n}}}$$

4.3. COMPARAREA DISPERSIILOR

Compararea a două dispersii. În cazul selecţiilor de volum mare, pentru stabilirea semnificaţiei diferenţei a două dispersii s_1^2 şi s_2^2, se calculează eroarea acestei diferenţe, după cum urmează:

$$s_d^2 = \sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}} \tag{4.18}$$

Ca şi în cazul precedent, semnificaţia diferenţelor se stabileşte în raport cu testul u:

$$u = \frac{s_1^2 - s_2^2}{s_d^2} \tag{4.19}$$

sau

$$u = \frac{s_1^2 - s_2^2}{\sqrt{\frac{s_1^2}{2n_1} + \frac{s_2^2}{2n_2}}} \tag{4.20}$$

În cazul *selecţiilor de volum mic*, în acest scop se recomandă criteriul FISHER:

$$F = \frac{s_1^2}{s_2^2} \tag{4.21}$$

Dacă s_2^2 este mai mare decît s_1^2 se studiază raportul:

$$F = \frac{s_2^2}{s_1^2} \tag{4.22}$$

În anexa 5 se prezintă distribuţia F pentru q=5%, 2,5% şi 1%. Valorile teoretice ale raportului F se citesc în funcţie de gradele de libertate f_1 şi f_2 şi de nivelul de semnificaţie ales q. Aceste valori se citesc la probabilitatea q:2.

Trebuie subliniat că în tabelul din anexa 5 valoarea criteriului F se caută la intersecţia coloanei corespunzătoare numărului gradelor de libertate ale dispersiei mai mari, cu rîndul corespunzător numărului gradelor de libertate ale dispersiei mai mici.

Scanned with OKEN Scanner

Exemplul 1. Studiind variația indicilor de formă la pinul silvestru din Carpații Orientali pe tipuri de pădure, Iacovlev* stabilește următoarele valori:

Tipul de pădure	n	$\bar{x}$	s	s^2
Pineta minora	120	0,703	0,085	0,0072
Pineta media	360	0,661	0,050	0,0025

Folosind formula 4.3 stabilim că diferența mediilor este distinct semnificativă, valoarea experimentală a testului u fiind de 2,5 (vezi § 4.2).

Pentru a stabili dacă diferența dispersiilor este semnificativă aplicăm formula (4.20):

$$u = \frac{0{,}0072 - 0{,}0025}{\sqrt{\frac{0{,}0072}{240} + \frac{0{,}0025}{720}}} = 8{,}1$$

Conchidem deci că diferența dintre cele două dispersii este foarte semnificativă, deoarece valoarea u experimentală este cu mult mai mare decît valoarea u teoretică. Prin urmare $\sigma_1^2 \neq \sigma_2^2$. Tipul de pădure *Pineta minora* se caracterizează printr-o amplitudine mare de variație a indicelui de formă. La tipul de pădure *Pineta media* forma arborilor este mai stabilă.

Exemplul 2. După datele din tabelul 4.2 obținem:

$$s_1^2 = 11{,}3;\ s_2^2 = 23{,}7:\ f_1 = 11;\ f_2 = 11$$

$$F = \frac{s_2^2}{s_1^2} = \frac{23{,}7}{11{,}3} = 2{,}1$$

După datele din anexa 5 stabilim că F teoretic este de 3,4 pentru q =5% (se citește valoarea corespunzătoare lui q: 2, respectiv la 2,5%). Rezultă deci că nu avem nici un temei să respingem ipoteza nulă, deoarece F experimental este mai mic decît F teoretic pentru q=5%.

În baza acestei concluzii, în exemplul 2 (§ 4.2) s-a folosit formula 4.5.

Exemplul 3. După datele din tabelul 4.3 rezultă:

$$s_1^2 = 0{,}53;\ s_2^2 = 0{,}10;\ f_1 = 11;\ f_2 = 11$$

$$F = \frac{0{,}53}{0{,}10} = 5{,}3$$

În tabelul din anexa 5, pentru nivelul de semnificație de 0,05, găsim F=3.4 (s-a citit la q:2=2,5%).

* Revista Pădurilor, nr. 2, 1961.

Scanned with OKEN Scanner

În baza acestor calcule conchidem că cele două dispersii diferă semnificativ.

Verificarea ipotezei omogenității mai multor dispersii. Omogenitatea dispersiilor de selecție se poate verifica cu ajutorul testului propus de BARTLETT. Dacă avem K selecții (variante) cu efectivul n_i și dispersiile s_i^2, se calculează inițial $s^2 = \Sigma \frac{f_i s_i^2}{f}$, unde $f_i = n_i - 1$ și $f = \Sigma f_i$. Apoi se determină mărimea $\chi^2 = \frac{1}{c}(\lg s^2 - \Sigma f_i \lg s_i^2)$, unde $c = 0{,}434 \left[1 + \frac{1}{3(K-1)}\left(\Sigma \frac{1}{f_i} - \frac{1}{f}\right)\right]$.

Valoarea χ^2 calculată se compară cu valoarea tabelară (anexa 6), folosind $K-1$ grade de libertate. Dacă χ^2 tabelar, la nivelul de semnificație ales, este mai mare decît χ^2 calculat, se acceptă ipoteza omogenității dispersiilor luate în considerare. Această verificare este utilă la analiza dispersională (cap.5).

Exemplu. Într-un arboret echien de gorun s-au făcut inventarieri parțiale prin 3 procedee. În total s-au amplasat 43 suprafețe de probă pentru fiecare variantă. După prelucrarea măsurătorilor efectuate pentru dispersia volumelor, s-au obținut următoarele valori:

procedeul 1 $s_1^2 = 0{,}39$ m³
procedeul 2 $s_2^2 = 0{,}41$ m³
procedeul 3 $s_3^2 = 0{,}38$ m³
1,18 m³

Pentru acest exemplu $s^2 = 1{,}18 : 3 = 0{,}393$, iar $f = 3(43-1) = 126$.

$$c = 0{,}434 \left[1 + \frac{1}{3 \times 2}\left(3 \frac{1}{42} - \frac{1}{126}\right)\right] = 0{,}44$$

$$\chi^2 = \frac{1}{0{,}44}(126 \lg 0{,}393 - 42 \lg 0{,}39 - 42 \lg 0{,}41 - 42 \lg 0{,}38) = 0{,}7$$

Valoarea tabelară, la nivelul de semnificație de 0,05 și la 2 grade de libertate, este de 5,99 (vezi anexa 6).

Întrucît valoarea tabelară a textului χ^2 este mai mare decît valoarea calculată (0,7) nu avem dovezi pentru a respinge ipoteza omogenității dispersiilor respective.

4.4. VERIFICAREA CONCORDANȚEI DINTRE DISTRIBUȚIA EMPIRICA ȘI DISTRIBUȚIA TEORETICA

În exemplele prezentate în capitolul 3 am presupus (am emis ipoteza) că distribuția caracteristicilor studiate urmează anumite legi de distribuție teoretică (binomială, POISSON, CHARLIER ș.a.).

Vom verifica acum în ce măsură datele experimentale confirmă sau nu ipoteza emisă, adică dacă există sau nu corespondență între distribuția empirică (experimentală) și distribuție teoretică.

După KOLGOMAROV verificarea corespondenței dintre distribuția empirică și distribuția teoretică se realizează astfel:

— se calculează frecvențele relative cumulate atît pentru distribuția empirică, cît și pentru distribuția teoretică;

Scanned with OKEN Scanner

— din frecvențele relative cumulate ale distribuției empirice se scad frecvențele relative cumulate ale distribuției teoretice;

— se stabilește cea mai mare diferență dintre cele două serii; se notează această diferență cu D;

— se calculează valoarea λ:

$$\lambda = D\sqrt{n} \tag{4.24}$$

— după datele din tabelul 4.5, în funcție de λ, se determină probabilitatea $1-K(\lambda)$;

— dacă $1-K(\lambda)$ este mai mare decît 0,05 (pentru un nivel de semnificație de 0,05, se conchide că variabila dată urmează legea de distribuție teoretică considerată*;

— dacă $1-K(\lambda)$ este mai mică deeît 0,05 nu există temei să se accepte ipoteza că variabila dată urmează legea de distribuție respectivă.

Tabelul 4.5

Valorile $1-K(\lambda)$ (după Mitropolski, *1961*)

λ	$1-K(\lambda)$	λ	$1-K(\lambda)$
0,3	0,99999	1,4	0,03968
0,4	99719	1,5	02222
0,5	96394	1,6	01195
0,6	86428	1 7	00618
0,7	71124	1,8	00307
0,8	54414	1,9	00146
0,9	39273	2,0	00067
1,0	27000	2,1	00030
1,1	17772	2,2	00013
1,2	11225	2,3	00005
1,3	06809	2,4	00002

E x e m p l u. Pentru exemplificare prezentăm în tabelul 4.6. frecvențele absolute experimentale (coloana 2) și frecvențele absolute teoretice (coloana 3) după datele din tabelul 3.7. La ajustarea distribuției empirice s-a adoptat ipoteza distribuției normale (formula 3.12). Se va verifica acum cu ajutorul criteriului lui KOLGOMOROV dacă ipoteza făcută este justă.

Tabelul 4.6

Verificarea concordanței dintre distribuția teoretică și distribuția experimentală pentru indicii de formă q_2 la molidul din țara noastră (n = 947)

q_2	Frecvențe relative		Frecvențe relative cumulate		diferențe
	empirice	teoretice	empirice	teoretice	
0,51	0,007	0,006	0,007	0,006	+ 0,001
0,54	0,024	0,020	0,031	0,026	+ 0,005
0,57	0,059	0,053	0,090	0,079	+ 0,011
0,60	0,101	0,108	0,191	0,187	+ 0,004
0,63	0,153	0,170	0,344	0,357	— *0,013*
0,66	0,221	0,205	0,565	0,562	+ 0,003
0,69	0,185	0,190	0,750	0,752	— 0,002
0,72	0,139	0,134	0,889	0,886	+ 0,003
0,75	0,077	0,072	0,966	0,958	+ 0,008
0,78	0,025	0,030	0,991	0,988	+ 0,003
0,81	0,005	0,010	0,996	0,998	— 0,002
0,84	0,004	0,002	1,000	1,000	0
Σ	1,000	1,000	—	—	—

* Pentru un nivel de semnificație de 0,01 ipoteza este acceptată numai dacă $1-K(\lambda)$ este mai mare de 0,01.

Scanned with OKEN Scanner

Din tabelul 4.6 rezultă: $D=0{,}013$.

După formula (4.24) obținem: $\lambda = 0{,}013.\sqrt{943} = 0{,}40$.

Pentru $\lambda=0{,}40$, după tabelul 4.5 obținem $1-K(\lambda)=0{,}997$.

Întrucît $1-K(\lambda)$ este cu mult mai mare decît 0,05 și foarte apropiată de 1,0, se conchide că distribuția arborilor de molid în raport cu indicile de formă q_2 urmează legea distribuției normale, potrivit formulei (3.12).

Tot prin criteriul KOLGOMOROV s-a mai stabilit că aceeași lege este urmată și de repartiția arborilor de molid în raport cu valorile indicilor de formă reali K_i (fig. 3.13).

Verificarea concordanței dintre distribuția empirică și distribuția teoretică se poate efectua și cu ajutorul testului PEARSON χ^2. (R a n c u-T ö v i s s i, 1963; M i h o c — U r s e a n u, 1962 ș.a.). După același criteriu se pot compara două distribuții experimentale.

* * *

Procedeele prezentate privind verificarea ipotezelor statistice se bazează pe un număr de valori observate, stabilite anterior, permițînd acceptarea sau respingerea ipotezei respective. W a l d (1947) propune însă procedeul *analizei secvențiale*, potrivit căruia observațiile se realizează sub forma unei serii succesive (T r e b i c i-R a n c u, 1964; L o e t s c h-H a l l e r, 1963).

După fiecare unitate observată (extrasă) se poate hotărî asupra: a) acceptării ipotezei; b) respingerii ipotezei; c) continuării experimentării. În cazul distribuțiilor normale această analiză se realizează ușor construind un grafic (fig. 4.2) cu următoarele două drepte:

$$T_o = a_o + bm \tag{4.25}$$

$$T_1 = a_1 + bm \tag{4.26}$$

unde:

$$a_o = 2{,}3026 \frac{\sigma^2}{\bar{x}_1 - \bar{x}_0} \lg \frac{\beta}{1-\alpha} \tag{4.27}$$

$$a_1 = 2{,}3026 \frac{\sigma^2}{\bar{x}_1 - \bar{x}_0} lg \frac{1-\beta}{\alpha} \tag{4.28}$$

$$b = \frac{\bar{x}_0 + \bar{x}_1}{2}. \tag{4.29}$$

În aceste formule semnificația simbolurilor este următoarea:

$\bar{x}_0$ — media variabilei x în prima ipoteză;
$\bar{x}_1$ — media variabilei x în a doua ipoteză $(\bar{x}_1 > \bar{x}_0)$;
α — probabilitatea acceptării ipotezei $\bar{x} = \bar{x}_1$, cînd în realitate $\bar{x} = \bar{x}_0$ (de obicei α se consideră egală cu 0,01 — 0,05);
β — probabilitatea acceptării ipotezei $\bar{x} = \bar{x}_0$, cînd în realitate $\bar{x} = \bar{x}_1$ (pentru β se aleg valori între 0,01 și 0,05);
m — numărul succesiv al unităților extrase;
σ — abaterea medie pătratică a distribuției respective. (Aceasta se presupune cunoscută din cercetări similare efectuate anterior).

Scanned with OKEN Scanner

Se dă: $\bar{X}_0 = 24{,}25$; $\bar{X}_1 = 24{,}50$

Alegem: $\alpha = 0{,}05$; $\beta = 0{,}05$

Din cercetări anterioare se cunoaște $\sigma = 0{,}25$

De unde:

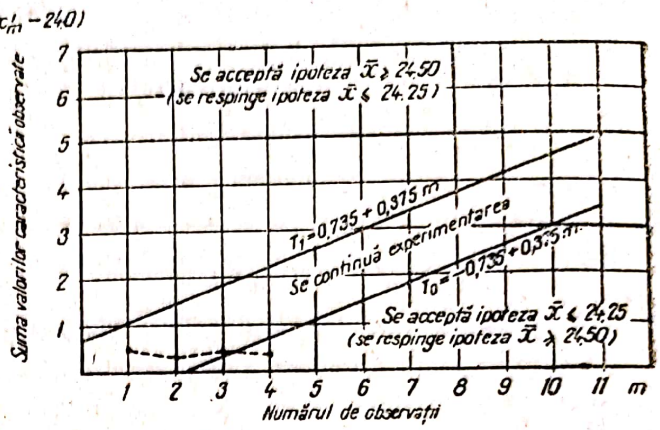
$$a_0 = \frac{2{,}3026 - 0{,}25^2}{24{,}50 - 24{,}25} \lg \frac{0{,}05}{1-0{,}05} = -0{,}735$$

$$a_1 = \frac{2{,}3026 - 0{,}25^2}{24{,}50 - 24{,}25} \lg \frac{1-0{,}05}{0{,}05} = +0{,}735$$

$$b = \frac{24{,}25 + 24{,}50}{2} = 24{,}375$$

$T_0 = -0{,}735 + 24{,}375$ m.

$T_1 = 0{,}735 + 24{,}375$ m



Rezultatele experimentării

Numărul curent al observației m	Valoarea caracteristicii x_m	Pentru simplificare introducem $x' = x_m - 24{,}0$	Valori cumulate $\sum (x_m - 24{,}0)$
1	24,50	0,50	0,50
2	23,78	−0,22	0,28
3	24,12	0,12	0,40
4	23,90	−0,10	0,30

După a patra observație avem temei pentru respingerea ipotezei $\bar{x} \geq 24{,}50$ respectiv pentru acceptarea ipotezei $\bar{x} \leq 24{,}25$

În raport cu noua variabilă x'_m obținem

$T_0 = -0{,}735 + 0{,}375$

$T_1 = 0{,}735 + 0{,}375$

Fig. 4.2. Un exemplu simplu de analiză secvențială

La verificarea ipotezei se procedează astfel:

— se continuă extracțiile atunci cînd punctele experimentale (reprezentînd suma valorilor caracteristicii observate) sînt cuprinse între cele două drepte T_0 și T_1 (fig. 4.2);

— dacă după extracția m_i suma valorilor caracteristicii corespunzătoare unităților extrase este un număr mai mic decît T_0, se acceptă ipoteza $\bar{x} \leqslant \bar{x}_0$ (în acest caz punctul experimental cu ordonata T_0 și abscisa m_i se situează sub linia T_0 (fig. 4.2);

— dacă după extracția m_i suma valorilor caracteristicii corespunzătoare unităților observate este un număr mai mare decît T_1 se acceptă ipoteza $\bar{x} \geqslant \bar{x}_1$ (punctul experimental cu ordonata T_1 și abscisa m_i se situează mai sus de linia T_1 (fig. 4.2).

*

Urmărind cunoașterea semnificației unor diferențe, permanent trebuie să avem în vedere următorul aspect deosebit de important: găsirea unei diferențe nesemnificative nu trebuie să ducă la concluzia că factorul cercetat are o influență redusă. Ea înseamnă numai că materialul de care dispunem este insuficient pentru a argumenta în mod temeinic o ipoteză; arată deci că materialul studiat nu este suficient de concludent. Această latură a problemei este ușor de înțeles dacă analizăm formula (4.2), de unde rezultă că erorile diferenței, deci și valorile u, depind de n (numărul repetițiilor), așa că neconcludența unor diferențe se poate datora unor selecții

Scanned with OKEN Scanner

cu un volum prea mic. În alte cazuri, diferențele pot fi nesemnificative din cauza organizării defectuoase a cercetărilor, mai ales atunci cînd au rămas neizolați anumiți factori perturbanți.

„Un rezultat neconcludent nu înseamnă o infirmare; el poate obliga însă la adîncirea unei cercetări. Invers, un rezultat statistic este asigurat nu înseamnă o concluzie definitivă. Asigurarea statistică înseamnă numai o probabilitate mică a ipotezei contrarii, în condiții de experiență care nu sînt niciodată total controlate" (S ă h l e a n u, 1957).

Scanned with OKEN Scanner

V

ANALIZA DISPERSIONALĂ CU APLICAȚII ÎN SILVICULTURĂ

5.1. GENERALITĂȚI

În majoritatea cercetărilor, mai ales în cele din domeniul silvobiologiei și al silvotehnicii, acționează simultan mai mulți factori, care generează o fluctuație a mărimilor statistice urmărite. O parte din această fluctuație se datorește influenței metodelor aplicate sau a factorilor experimentați; altă parte a variației poate fi însă cauzată de factori secundari neluați în considerare la organizarea experimentării; în sfîrșit, asupra rezultatelor pot influența și erorile de măsurare. De exemplu, într-o cercetare privind stabilirea influenței anumitor îngrășăminte asupra dezvoltării puieților în pepinieră, se obțin rezultate ce diferă uneori sensibil unul față de altul. Această fluctuație este cauzată de: variația datorită îngrășămîntului aplicat în diferite doze; variația cauzată de diferențe microstaționale sau de diversitatea ereditară a semințelor folosite; variația datorită erorilor de măsurare.

Analiza dispersională (analiza dispersiei, analiza variației, analiza împrăștierii) permite o valorificare superioară a datelor experimentale prin separarea (scindarea) fluctuațiilor rezultatelor experimentale în categorii de variații. Devine astfel posibilă evidențierea — pe baze științifice — a aportului fiecărei cauze la fluctuația generală.

La analiza dispersională un rol important îl are testul F (vezi § 4.3). Dacă valorile experimentale ale testului F sînt mai mici decît valorile teoretice (tabelul din anexa 6) în funcție de gradele de libertate corespunzătoare, atunci nu avem nici un temei să afirmăm că factorul studiat influențează asupra mărimii statistice respective; din contră, dacă F experimental este mai mare decît F tabelar, atunci trebuie acceptată ipoteza că factorul cercetat are o influență semnificativă.

Dar, ca și în cazul testului t, folosit la stabilirea semnificației diferențelor dintre medii, deosebirile nesemnificative rezultate din analiza dispersională efectuată nu trebuie confundate cu lipsa existenței unor eventuale deosebiri. Rezultatul neconcludent poate obliga însă la aprofundarea cercetărilor prin majorarea numărului de repetiții.

Scanned with OKEN Scanner

După ce, prin intermediul analizei dispersionale s-au constatat diferențe semnificative (ceea ce înseamnă că cel puțin două variante diferă între ele semnificativ) se trece la o analiză de detaliu a influenței diferiților factori luați în considerare. În acest scop, cu ajutorul testului t se stabilește între ce anume variante există diferențe semnificative. În multe situații sînt mai indicate testele mai severe propuse de DUNKAN, TUKEY, NEWMAN-KEULS (S n e d e c o r, 1957).

Desfășurarea calculelor în cadrul analizei dispersionale depinde în mare măsură de modul cum s-au organizat inițial lucrările experimentale. Important de remarcat este legătura strînsă dintre planificarea și organizarea experimentărilor, pe de o parte, și prelucrarea statistico-matematică a rezultatelor obținute pe de altă parte. De obicei, pentru un anumit mod de organizare a experimentărilor există o singură metodă corespunzătoare pentru prelucrarea și analiza rezultatelor. Alegerea unei scheme nereușite poate duce la calcule complicate și uneori mai puțin concludente; altă dată este posibil ca rezultatele obținute să nu răspundă la anumite întrebări importante. De aceea, trebuie alese scheme profund analizate și din punct de vedere al viitoarei prelucrări statistice a rezultatelor obținute. Numai astfel evităm eventuale surprize care pot să apară în faza de prelucrare a datelor experimentale. Experiențele defectuos organizate nu pot fi eficient valorificate prin prelucrări ulterioare statistico-matematice.

5.2. ANALIZA DISPERSIONALĂ IN CAZUL UNEI GRUPĂRI SIMPLE

Tehnica analizei dispersionale depinde în mare măsură de numărul cauzelor generatoare de fluctuații luate în considerare la organizarea experimentărilor. Dacă variantele care generează fluctuația rezultatelor se referă la aceeași cauză, atunci, în scopul prelucrării datelor, se formează o grupare simplă. De exemplu, urmărim să stabilim influența lungimii manelelor de răşinoase cu diametrul de 10 cm la capătul subțire asupra coeficientului de echivalență a numărului de metri liniari în metri cubi. În acest caz s-a luat în considerare un singur factor — lungimea pieselor (cu patru variante: 3, 4, 5 și 6 m)

Datele experimentale se trec într-un tabel de forma celui prezentat mai jos (tabelul 5.1).

Pentru a putea compara mediile $\bar{x}_i$ obținute pe variante, acceptăm ipoteza potrivit căreia toate mediile $\bar{x}_i$ sînt egale între ele, adică sînt estimații ale aceleiași medii generale. S-a arătat că unei astfel de ipoteze i se spune *ipoteză nulă.* Pentru a accepta sau respinge ipoteza nulă va trebui să calculăm raportul a două dispersii.

Mărimile statistice necesare se calculează după cum urmează:

În primul rînd se determină *suma totală a pătratelor* abaterilor de la media generală $\bar{x}$. Obținem astfel expresia:

$$\sum (x_{ij} - \bar{x})^2 = \sum x_{ij}^2 - \frac{G^2}{N} = Q^* \qquad (5.1)$$

*) În unele publicații se folosește simbolul SP.

Scanned with OKEN Scanner

Tabelul 5.1

Înregistrarea datelor experimentale în cazul unei grupări simple

Variantele factorului A	Valori observate x_{ij}	Numărul valorilor observate	Suma valorilor T_i observate	Media $\bar{x}_i$
1	$x_{11}, x_{12}, \ldots, x_{1n_1}$	n_1	T_1	$\bar{x}_1$
2	$x_{21}, x_{22}, \ldots, x_{2n_2}$	n_2	T_2	$\bar{x}_2$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
i	$x_{i1}, x_{i2} \ldots, x_{in_i}$	n_i	T_i	$\bar{x}_i$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
K	$x_{k1}, x_{k2}, \ldots, x_{kn_k}$	n_k	T_k	$\bar{x}_k$
	Total	$N = \sum n_i$	$G = \sum T_i$	$\bar{x} = \frac{G}{N}$

sau înlocuind

$$\frac{G^2}{N} = C \tag{5.2}$$

rezultă expresia

$$Q = \sum x_{ij}^2 - C. \tag{5.3}$$

În continuare calculăm suma pătratelor între variante:

$$Q_A = \frac{\sum T_i^2}{n_i} - C \tag{5.4}$$

unde însumarea se efectuează pentru toate variantele.

Rezultatele obținute se înscriu într-un tabel standard (tabelul 5.2).

Tabelul 5.2

Schema analizei dispersionale pentru grupări simple

Împrăștierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul F
Intre variante	$Q_A = \frac{\sum T_i^2}{n_i} - C$	$K-1$	$s_A^2 = \frac{Q_A}{K-1}$	$F = \frac{s_A^2}{s_R^2}$
In interiorul variantelor (reziduală)	$Q_R = Q - Q_A$	$N-K$	$s_R^2 - \frac{Q_R}{N-K}$	
Totală	$Q = \sum x_{ij}^2 - C$	$N-1$	$s^2 = \frac{Q}{N-1}$	

Scanned with OKEN Scanner

Mărimea Q_R, care exprimă *suma pătratelor în interiorul variantelor*, se obține prin scădere:

$$Q_R = Q - Q_A \tag{5.5}$$

Pentru control se poate folosi formula:

$$Q_R = \sum_{i=1}^{K} \left\{ \sum_{j=1}^{n_i} x_{ij}^2 - \frac{T_i^2}{n_i} \right\} \tag{5.6}$$

În continuare se trece la determinarea numărului gradelor de libertate. Suma totală a pătratelor abaterilor de la media generală are $f=N-1$ grade de libertate. Suma pătratelor abaterilor privind împrăștierea între variante prezintă $f_1=K-1$ grade de libertate. Prin scădere obținem numărul gradelor de libertate pentru suma reziduală: $f_2=(N-1)-(K-1)=N-K$.

În baza celor stabilite calculăm dispersiile s_A^2, s_R^2 și s^2 (tabelul 5.2).

Dacă s_A este semnificativ mai mare decît s_R, atunci se respinge ipoteza nulă și acceptăm că între variante există diferențe reale. Dacă s_A este semnificativ mai mică decît s_R, atunci acestui fapt nu i se acordă o importanță deosebită, dar în cazul cînd acest rezultat se repetă, trebuie luat în considerare. La această analiză, deosebit de util este testul F

$$F = \frac{s_A^2}{s_R^2} \tag{57.}$$

Valoarea experimentală a testului F se compară cu valoarea tabelară folosind distribuția F (anexa 5), în raport cu numărul gradelor de libertate f_1 și f_2. Ipoteza nulă se respinge atunci cînd F calculat este mai mare decît F tabelar. În caz contrar nu avem nici un temei pentru respingerea ipotezei nule.

Constatînd prin criteriul F că există diferențe între variantele studiate, în continuare se trece la examinarea asigurării statistice a diferențelor puse în lumină. În acest scop, după cum s-a mai arătat, folosim criteriul u ori t sau alte criterii mai severe, cum sînt cele propuse de DUNCAN, NEWMAN-KEULS ș.a.

Exemplul 1. La organizarea unor experimentări s-a presupus că lungimea manelelor de rășinoase (cu diametrul de 10 cm la capătul subțire) influențează asupra coeficientului de echivalență a metrilor liniari în metri cubi. Măsurătorile efectuate la 44 loturi experimentale pentru patru lungimi (4 variante ale factorului A) sînt prezentate în tabelul 5.3. Loturile experimentale au fost alese la întîmplare.

Calculele se desfășoară după cum urmează:

$$N = 44;\ G = \Sigma x_{ij} = 4129;\ C = \frac{G^2}{N} = \frac{4129^2}{44} = 387\,469$$

Suma totală a pătratelor abaterilor de la media sistemului ($\bar{x}=94$) va fi:

$$Q = \Sigma x_{ij}^2 - C = (100^2 + 95^2 + 107^2 + \ldots + 88^2) - 387\,469 = 3748$$

Scanned with OKEN Scanner

Tabelul 5.3

Lungimi A_i	Valori observate x_{ij}	Numărul valorilor observate n_i	Suma T_i	Media x_i
3	100, 95, 107, 102, 105, 93	6	602	100
4	103, 95, 101, 96, 98, 98 114 92, 92, 95, 95, 93, 100, 98	14	1 370	98
5	97, 96, 92, 99, 90, 89, 75, 116, 92, 82, 96, 95, 97	13	1 218	94
6	95, 80, 79, 93, 82, 64, 89, 86 97, 88, 88,	11	941	86
	Total	44	4 129	94

Suma pătratelor corespunzătoare diferențelor dintre variante (lungimi) reprezintă:

$$Q_A = \sum \frac{T_i^2}{n_i} - C = \left(\frac{602^2}{6} + \frac{1370^2}{14} + \frac{1218^2}{13} + \frac{941^2}{11}\right) - 387\,469 = 1611$$

În sfîrșit, trecem la completarea tabelului standard pentru analiza dispersională (tabelul 5.4)

Tabelul 5.4

Tabelul pentru analiza dispersională, după datele tabelului *5.3*

Împrăștierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia s^2	Testul F
Intre variante	$Q_A = 1\,611$	3	537,0	
In interiorul variantelor (reziduală)	$Q_R = 2\,137$	40	53,0	10,1
Totală	$Q = 3\,748$	43	—	

Între variante avem $f=4-1=3$ grade de libertate. Pentru dispersia reziduală stabilim $f=44-4=40$ grade de libertate. Din tabel se vede clar că dispersia între variante, determinată de factorul luat în considerare, este cu mult mai mare decît dispersia reziduală, generată de factori neluați în seamă. Raportul $F = \frac{s_A^2}{s_R^2} = 10{,}1$. După tabelul din anexa 5 la $f_1=3$ și $f_2=40$ grade de libertate găsim o valoare a testului F de 2,9, la nivelul de semnificație de 5%. Pentru nivelul de semnificație de 1%, F tabelar este de 4,3.

Scanned with OKEN Scanner

Prin urmare, testul F dezvăluie o mare neuniformitate între variante, deoarece F tabelar, pentru $q=1\%$, este inferior raportului F experimental. Se demonstrează deci că aceste diferențe semnalate între variante nu se datoresc unor cauze întîmplătoare și că ele apar ca un efect evident al factorului A (lungimea manelelor). De aici concluzia practică: lungimea pieselor trebuie necondiționat luată în considerare la stabilirea indicilor de echivalență a metrilor liniari în metri cubi.

Admițînd ipoteza că dispersiile de selecție sînt estimații ale aceleași dispersii generale, în baza dispersiei s_R^2, se determină erorile mediilor făcînd raportul

$$\frac{s_R}{\sqrt{n_i}} \qquad (5.8)$$

Tabelul 5.5

Lungimi, m (variante)	media $\bar{x}_i$	eroarea medie $\frac{s_R}{\sqrt{n_i}}$
3	100	$\pm 3{,}0$
4	98	$\pm 1{,}9$
5	94	$\pm 2{,}0$
6	86	$\pm 2{,}2$

Pentru exemplul considerat $s_R^2=53{,}0$ de unde $s_R=7{,}3$. Prin urmare, cele patru medii, pe variante, vor avea erorile menționate în tabelul 5.5.

În continuare se trece la stabilirea gradului de semnificație a diferențelor. În acest scop se folosește testul t, împărțind fiecare diferență respectivă (între fiecare variantă și media variantelor sau între fiecare variantă și o variantă martor, sau între două variante) la eroarea diferențelor:

$$t = \frac{\bar{x}_i - \bar{x}_j}{s_R\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}} \qquad (5.9)$$

Valoarea t astfel calculată se compară cu valoarea t tabelară, folosind, în raport cu numărul valorilor observate, fie distribuția normală (anexa 3), fie distribuția S t u d e n t pentru $N - K$ grade de libertate (anexa 4).

Astfel, pentru prima variantă ($l=3$ m) și ultima variantă ($l=6$ m) obținem:

$$t = \frac{100-86}{7{,}3\sqrt{\frac{1}{6}+\frac{1}{11}}} = 3{,}8.$$

(A se vedea și tabelul 5.3).

Potrivit distribuției S t u d e n t (anexa 4), la nivelul de semnificație 0,05 și la $f=N-K=44-4=40$ grade de libertate, rezultă $t=2{,}02$. Valoarea calculată a testului t fiind mai mare decît valoarea tabelară conchidem că diferența dintre cele două variante se datorește unor cauze reale.

Semnificația diferențelor se mai poate stabili cu ajutorul tabelei din anexa 7. Cunoscînd numărul gradelor de libertate al dispersiei s_R^2, respectiv $N - K$, se caută în tabelul din anexa 7 pentru fiecare valoare t experimentală (calculată), probabilitatea de transgresiune corespunzătoare fiecăre-

Scanned with OKEN Scanner

diferențe. Astfel, pentru cele două variante, la 40 grade de libertate și pentru t calculat de 4,6, rezultă o probabilitate de transgresiune mai mică decît 0,1%, ceea ce arată că diferența dintre cele două medii este foarte semnificativă.

În acest fel se repetă calculul și pentru alte variante.

Exemplul 2. Într-un arboret de pin (70 ani) la 236 arbori s-au efectuat măsurători de creștere cu ajutorul burghiului PRESSLER, luînd probe din patru părți (sud, est, nord și vest) ale secțiunii de bază. Rezultatele obținute sînt centralizate în tabelul 5.6. Ne propunem să vedem dacă

Tabelul 5.6

Creșterile pe rază la *236* arbori, măsurate din patru puncte ale secțiunii de bază (arboret de pin, *70* ani)*

Varianta	Creșterea pe rază, mm									Total	Suma T_i	Media x_i
	5	8	11	14	17	20	23	26	29			
	frecvența											
Sud	12	39	53	62	39	11	8	6	6	236	3 220	13,64
Est	12	28	32	54	44	32	16	10	8	236	3 640	15,42
Nord	12	32	42	54	34	30	18	9	5	236	3 505	14,85
Vest	12	48	52	50	38	20	6	5	5	236	3 175	13,45
Total	48	147	179	220	155	93	48	30	24	944	13 540	14,34

* Extras din lucrarea „Determinarea creșterii curente la arborete" (Giurgiu, 1957)

mediile de selecție obținute pentru cele patru puncte cardinale au valori diferite datorită influențelor sistematice ale poziției locului de unde s-a extras proba (față de punctele cardinale) sau dacă valorile medii sînt diferite numai datorită influenței factorilor întîmplători.

În acest scop calculăm:

$$N = 944$$

$$G = \Sigma x_i = 5 \times 48 + 8 \times 147 + \ldots + 29 \times 24 = 13\,540$$

$$C = \frac{G^2}{N} = \frac{13\,574^2}{944} = 194\,207$$

Suma totală a pătratelor abaterilor de la media sistemului ($\bar{x}$=14,34) va fi:

$$Q = \sum x_{ij}^2 - C = [(5^2 \times 48 + 8^2 \times 147 + \ldots + 29^2 \times 24)] - 194\,207 = 29\,031.$$

Suma pătratelor corespunzătoare diferențelor dintre variante (puncte cardinale) reprezintă:

$$Q_A = \sum \frac{T_i^2}{n_i} - C = \left[\frac{3\,220^2}{236} + \frac{3\,640^2}{236} + \frac{3\,505^2}{236} + \frac{3\,175^2}{236}\right] - 194\,207 = 639$$

Suma pătratelor pentru împrăștierea reziduală rezultă prin diferență:

$$Q_R = Q - Q_A = 29\,031 - 639 = 28\,392.$$

Scanned with OKEN Scanner

Introducînd rezultatele obținute în tabelul analizei dispersionale obținem (tabelul 5.6 *a*):

$$F = \frac{213}{30} = 7{,}1$$

F tabelar (anexa 5), la nivelul de semnificație de 0,05, la $f_1 = 3$ și $f_2 = 940$ grade de libertate, este de 2,6.

Cum *F* calculat este mai mare decît *F* tabelar (7,1 > 2,6) se consideră că mediile de selecție nu sînt estimații ale aceleeași medii a colectivității generale și că rezultatele diferă semnificativ, tocmai datorită influenței reale a poziției față de punctele cardinale. Stabilim că diferențele dintre variante se datoresc unor influențe reale și nu unor factori întîmplători.

Eroarea mediilor va fi $\sqrt{30}/\sqrt{n_i} = 5{,}46\sqrt{n_i}$.

Tabelul 5.6 a

Tabelul pentru analiza dispersională, după datele tabelului *5.6*

Împrăștierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia s^2	Testul *F*
Intre variante	639	3	213	213 : 30 = 7,1
In interiorul variantelor (reziduală)	28 392	940	30	
Totală	29 031	943	—	—

În continuare se stabilește semnificația diferențelor dintre variante. Astfel, pentru diferența dintre primele două variante obținem:

$$u = \frac{13{,}64 - 15{,}42}{5{,}46\sqrt{\frac{1}{236} + \frac{2}{236}}} = 3{,}55$$

După anexa 3 *), la $u = 3{,}55$, corespunde o probabilitate de 0,999 (probabilitatea de depășire fiind deci de 0,001), ceea ce dovedește că diferența dintre cele două variante este foarte semnificativă. În mod similar se pot repeta calculele pentru restul variantelor.

Compararea mediilor prin metoda DUNCAN Cu mult mai riguros compararea mediilor se efectuează după metoda DUNCAN elaborată în anul 1955 (H u c h s, 1963). În esență, se procedează astfel (după datele exemplului 2):

a) Prin analiza dispersională se verifică ipoteza nulă, folosind testul *F*.

b) În cazul respingerii ipotezei nule se calculează eroarea standard a mediilor după formula cunoscută:

$$s_{\bar{x}} = \sqrt{s_R^2/n} = \sqrt{30/236} = 0{,}36$$

unde *n* reprezintă numărul repetițiilor.

* Se folosește anexa 3 (distribuția normală) deoarece numărul valorilor observate este ridicat.

c) Mediile se înscriu în ordine crescătoare, după cum urmează:

varianta	V	S	N	E
media, mm	13,45	13,64	14,85	15,42
rangul	1	2	3	4

d) După datele din anexa 8, la nivelul de semnificație ales, în funcție de numărul gradelor de libertate al împrăștierii reziduale ($f = N - K$) și de numărul mediilor cuprinse în serie (p), se determină diferențele semnificative tabelare (p are valori de la 2 la k).

Pentru exemplul considerat la nivelul de semnificație 0,05 obținem:

p	2	3	4
diferențe semnificative tabelare (anexa 8)	2,77	2,92	3,02

e) Se calculează *cele mai mici diferențe semnificative* multiplicînd ultimele valori tabelare cu eroarea standard $s_{\bar{x}}$:

p	2	3	4
cele mai mici diferențe semnificative	1,00	1,05	1,09

f) În continuare se calculează diferențele dintre medii, în următoarea ordine:

— din valoarea cea mai mare se scade cea mai mică;
— din valoarea cea mai mare se scade cea mai mică din cele rămase etc.;
— din a doua valoare ca mărime se scade cea mai mică;
— din a doua valoare ca mărime se scade cea mai mică din cele rămase (etc.)

g) Diferențele astfel determinate se consideră semnificative la nivelul de semnificație ales, atunci cînd sînt mai mari decît *cele mai mici diferențe semnificative* calculate la punctul *e*.

h) Rezultatele calculelor se sistematizează după cum urmează:

Se compară variantele (rangurile)	Diferențe	p*	Cele mai mici diferențe semnificative	Diferența dintre medii este:
4 cu 1 (E-V)	15,42 — 13,45 = 1,97	4	1,16	semnificativă
4 cu 2 (E-S)	15,42 — 13,64 = 1,78	3	1,14	semnificativă
4 cu 3 (E-N)	15,42 — 14,85 = 0,57	2	1,08	nesemnificativă
3 cu 1 (N-V)	14,85 — 13,45 = 1,40	3	1,14	semnificativă
3 cu 2 (N-S)	14,85 — 13,64 = 1,21	2	1,08	semnificativă
2 cu 1 (S-V)	13,64 — 13,45 = 0,19	2	1,08	nesemnificativă

* Pentru primul rînd $p = 4$, deoarece în serie sînt cuprinse 4 medii : 13,45; 13,64; 14,85; și 15,42. Pentru rîndul al treilea $p=2$, în serie fiind cuprinse numai două medii: 14,85 și 15,42.

Scanned with OKEN Scanner

Analiza dispersională a scos în evidenţă neomogenitatea sistemului, arătînd că cel puţin două medii diferă semnificativ una faţă de alta. Metoda DUNCAN a precizat care sînt combinaţiile de variante cu diferenţe semnificative între medii.

*

În cazul cînd valoarea calculată a lui F este mai mare decît valoarea tabelară F, analiza dispersională este uneori folosită în continuare pentru omogenizarea colectivităţii statistice, prin excluderea variantelor (mediilor) care se abat de la media generală. În acest scop, se exclude mai întîi varianta (media) cu cea mai mare abatere şi se verifică încă o dată omogenitatea seriilor statistice studiate. În acest mod se ajunge treptat la omogenizarea colectivităţii statistice luată în studiu, prin eliminarea variantelor (mediilor) care nu estimează media generală.

Este deosebit de util acest mod de analiză pentru verificarea omogenităţii anumitor arborete sub raportul caracteristicilor solului (de pildă după pH, profunzime, textură etc.) sau al caracteristicilor vegetaţiei forestiere. Tot aşa de bine poate fi folosit pentru verificarea omogenităţii seminţelor forestiere de proveniență diferită în raport cu indicii lor calitativi, precum şi în multe alte diverse situaţii.

O condiţie esenţială pentru aplicarea analizei dispersionale prin grupare simplă este aceea de a asigura un caracter strict întîmplător efecturăii observaţiilor.

Metoda descrisă prezintă unele avantaje şi dezavantaje, dintre care semnalăm următoarele:

a) Metoda poate fi folosită la experienţele cu un număr mare de variante şi cu un număr diferit de observaţii. În plus, numărul valorilor observate nu trebuie să fie egal pentru toate variantele. (Acest aspect prezintă o deosebită importanţă pentru experimentările ce se fac în silvicultură, unde cercetătorul de multe ori este obligat să organizeze cercetările după vegetaţia forestieră instalată în mod natural, neputînd astfel realiza un număr egal de valori observate pentru toate variantele alese).

b) Prelucrarea statistică a materialului observat este relativ simplă. Mai mult decît atît, rezultatele nu se schimbă dacă o parte sau toate măsurătorile unei anumite variante, din cauze accidentale, nu mai pot fi valorificate. (Pentru alte scheme de organizare a experimentărilor, pierderea unor observaţii obligă la o analiză specială a situaţiei).

c) La experienţele de cîmp metoda descrisă este inferioară altor metode, prin aceea că furnizează informaţii mai puţine. Prezintă dezavantaje acolo unde nu se întîlneşte o bună uniformitate a solului. Dezavantajul semnalat se înlătură parţial prin folosirea metodei descrise în paragraful următor.

În exemplele date şi în expunerea teoretică efectuată mai sus, s-a presupus că dispersia este egală pentru toate variantele. În realitate însă, această egalitate uneori nu se constată şi de aceea se recomandă o verificare prealabilă, folosind criteriul omogeneităţii dispersiilor (testul BARTLETT, § 4.3). În practică, uneori se renunţă însă la această verificare, menţinîndu-se numai atunci cînd există un dubiu în legătură cu omogenitatea dispersiilor luate în considerare.

Scanned with OKEN Scanner

5.3. ANALIZA DISPERSIONALA IN CAZUL A DOI FACTORI (GRUPAREA DUBLA)

Această metodă permite analiza influențelor exercitate de doi factori A și B, fiecare dintre aceștia avînd mai multe variante. În cazul experimentelor în cîmp (pepiniere, plantații forestiere etc.), variantele factorului B reprezintă blocurile experimentale instalate (vezi § 5.5).

Fie x o caracteristică — spre exemplu conținutul în celuloză la lemnul de pin — ale cărei valori depind de doi factori A (tipul de pădure) și B (vîrsta arboretului). Datele privind variația caracteristicii în raport cu cei doi factori se trec într-un tabel tip (tabelul 5.7).

Tabelul 5.7

Tabelul datelor primare pentru analiza dispersională în cazul a doi factori

Variantele factorului A	Variantele factorului B (blocul)					Suma	Media $\bar{x}_i$
	1	2	.....	j	b		
1	x_{11}	x_{12}	.....	x_{1j}	x_{1b}	T_1	$\bar{x}_1$
2	x_{21}	x_{22}	.....	x_{2j}	x_{2b}	T_2	$\bar{x}_2$
.	.	.			.	.	.
.	.	.			.	.	.
.	.	.			.	.	.
i	x_{i1}	x_{i2}	.....	x_{ij}	x_{ib}	T_i	$\bar{x}_i$
.	.	.			.	.	.
.	.	.			.	.	.
.	.	.			.	.	.
t	x_{t1}	x_{t2}	.....	x_{tj}	x_{tb}	T_t	$\bar{x}_t$
Suma	B_1	B_2	.....,	B_j	B_b	G	$\bar{x} = G/bt$

În tabelul 5.7 se prezintă cel mai simplu caz, cînd pentru fiecare combinație a factorului A cu B corespunde o singură valoare, respectiv s-a efectuat o singură observație. Pentru factorul B sînt b variante (sau b blocuri). Combinației variantei i a factorului A cu varianta j a factorului B îi corespunde valoarea x_{tj}. Suma valorilor pentru varianta i a factorului A este T_i, iar suma valorilor pentru varianta j a factorului B este B_j. Suma valorilor pentru toate combinațiile sistemului ($b.t$) este egală cu G.

În cazul de față avem:

$$C = \frac{G^2}{bt} \tag{5.10}$$

În tabelul 5.8 se prezintă schema analizei dispersionale pentru grupări duble, de unde rezultă modul de stabilire a numărului gradelor de libertate, precum și a dispersiilor necesare pentru determinarea testului F.

Scanned with OKEN Scanner

Tabelul 5.8

Schema analizei dispersionale pentru grupări duble (în cazul a doi factori)

Împrăştierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul F
Între variantele factorului A	$Q_A = \sum_i \frac{T_i^2}{b} - C$	$t-1$	$s_A^2 = \frac{Q_A}{t-1}$	$\frac{s_A^2}{s_R^2}$
Între variantele factorului B (sau între blocuri)	$Q_B = \sum_j \frac{B_j^2}{t} - C$	$b-1$	$s_B^2 = \frac{Q_B}{b-1}$	$\frac{s_B^2}{s_R^2}$
Reziduală	$Q_R = Q - Q_A - Q_B$	$(t-1)(b-1)$	$s_R^2 = \frac{Q_R}{(t-1)\ (b-1)}$	—
Totală	$Q = \sum x_{ij}^2 - C$	$bt-1$	—	—

În caz că se constată diferenţe în sistemul analizat, se trece la determinarea gradului de semnificaţie al diferenţelor, folosind testul t. Pentru diferenţele $\bar{x}_i - \bar{x}_j$ avem:

$$t = \frac{x_i - x_j}{s\sqrt{\frac{2}{b}}} \qquad (5.11)$$

Eroarea medie pentru media variantelor factorului A se calculează după formula:

$$s_{x_i} = \frac{s}{\sqrt{b}} \qquad (5.12)$$

În baza acestei erori se pot stabili limitele intervalului de încredere.

Exemplul 1. În tabelul 5.9 sînt trecute rezultatele unor cercetări privind conţinutul în celuloză la lemnul de pin în raport cu tipul de

Tabelul 5.9

Conţinutul în celuloză la lemnul de pin în raport cu tipul de pădure şi vîrsta arboretului

Tipul de pădure (factorul A)	Clasa de vîrstă (factorul B)						Suma	Media şi eroarea mediei $\bar{x}_i \pm s_{\bar{x}_i}$
	III	IV	V	VI	VII	VIII		
1	58	63	53	57	52	55	338	56,3 ± 0,85
2	57	59	57	57	54	54	338	56,3 ± 0,85
3	52	63	53	56	54	52	330	55,0 ± 0,85
4	53	59	50	55	50	50	317	52,8 ± 0,85
5	52	57	54	59	47	54	323	53,8 ± 0,85
Suma	274	301	267	284	257	265	1 646	54,9

Scanned with OKEN Scanner

pădure și vîrsta arboretului. Pentru fiecare combinație a factorului A cu factorul B s-a efectuat o singură observație.

După datele din tabelul 5.9 rezultă:

$$N = bt = 30$$

$$G = 1646$$

$$C = 2\,709\,316 : 30 = 90\,310$$

Suma totală a pătratelor abaterilor față de media 54,9 este:

$$Q = \Sigma x_{ij}^2 - C = 58^2 + 63^2 + 53^2 + 57^2 + \ldots + 54^2 - 90\,310 = 398$$

Suma pătratelor corespunzătoare factorului A:

$$\sum \frac{T_i^2}{b} - C = \frac{1}{6}(338^2 + 338^2 + 330^2 + 317^2 + 323^2) - 90\,310 = 58$$

Suma pătratelor pentru factorul B:

$$\sum \frac{B_j^2}{t} - C = \frac{1}{5}(272^2 + 301^2 + \ldots + 265^2) - 90\,310 = 251$$

În continuare, întocmim tabelul analizei dispersionale (tabelul 5.10).

Tabelul 5.10

Rezultatele analizei dispersionale, după datele din tabelul *5.9*

Împrăștierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul F
Între tipurile de pădure	58	4	14,5	3,31
Între vîrste	251	5	50,2	11,4
Reziduală	89	20	4,4	—
Totală	398	29	—	—

Valorile calculate ale testului F se compară cu valorile tabelare corespunzătoare gradelor de libertate (4; 20), (5; 20) și nivelului de semnificație de 5%. Din anexa 5 obținem:

$$F_{4;\,20} = 2{,}9;\quad F_{5;\,20} = 2{,}7$$

Se constată că valorile calculate sînt mai mari decît valorile tabelare corespunzătoare (se respinge ipoteza nulă), așa încît conchidem că cei doi factori luați în considerare (tipul de pădure și vîrsta arboretului) exercită influențe semnificative asupra conținutului în celuloză la lemnul de pin.

Eroarea mediilor pentru variantele factorului A este egală cu $\sqrt{s^2}/\sqrt{b} = \sqrt{4{,}4}/\sqrt{6} = 0{,}85$ (tabelul 5.9), iar eroarea diferențelor este egală cu $\sqrt{4{,}4} \times \sqrt{\frac{2}{6}} = 1{,}2$.

Semnificația diferențelor dintre medii se calculează prin intermediul testului t (formula 5.11 și tabelul din anexa 4, folosind 20 grade de libertate). Spre exemplu, între tipurile de pădure 1 și 4 obținem:

$$t = \frac{56{,}3 - 52{,}8}{1{,}2} = 3{,}3.$$

Valoarea tabelară a testului t, la nivelul de semnificație de 0,05 și la 20 grade de libertate este de 2,09. La nivelul de semnificație de 0,01 valoarea tabelară a testului t este de 2,85. Diferența este deci distinct semnificativă. La același rezultat ajungem și după tabelul din anexa 7. Rezultate mai riguroase se obțin prin metoda DUNCAN (§ 5.2).

O condiție esențială la aplicarea acestei metode este aceea ca să nu lipsească nici o combinație dintre variantele factorului A cu cele ale factorului B. Atunci cînd din anumite cauze pentru una din aceste combinații (x'_{ij}) nu s-au efectuat observații, este necesar un calcul teoretic folosind formula:

$$x'_{ij} = \frac{tT'_i + bB'_j - G'}{(t-1)(b-1)} \qquad (5.13)$$

unde T'_i, B'_j și G' — corespund valorilor observate în realitate. Apoi se efectuează analiza dispersională după metoda descrisă, luînd în considerare și valoarea calculată x'_{ij}, dar se scade cîte un grad de libertate pentru Q și Q_R. Dacă lipsesc mai multe repetiții se recomandă să se renunțe la varianta respectivă.

* * *

În exemplul precedent, pentru simplificare, s-a considerat o singură observație pentru fiecare combinație. Dar acest mod de organizare a experimentărilor nu permite în toate cazurile obținerea unor rezultate concludente. De aceea se recomandă folosirea mai multor observații pentru fiecare combinație, adică să efectuăm mai multe repetiții pentru fiecare x_{ij}. Totodată, prezintă interes și interacțiunea factorilor A și B.

E x e m p l u l 2. Să presupunem că într-o experiență privind influența tipului de pădure și a poziției arborelui în arboret asupra rezistenței la forfecare (perpendicular pe fibre) a lemnului de molid s-au efectuat patru măsurători la fiecare combinație a factorului A (tipul de pădure) cu factorul B (poziția arborelui). Rezultatele obținute sînt prezentate în tabelul 5.11.

Pentru simplificarea calculelor, toate valorile observate au fost micșorate cu 300 (ceea ce nu modifică cu nimic concluziile analizei dispersionale). Datele astfel micșorate sînt prezentate în tabelul 5.12. În tabelul 5.13 sînt însumate toate valorile observate, astfel micșorate, iar în tabelul 5.14 aceste valori sînt ridicate la pătrat.

Cu ajutorul acestor date trecem la determinarea elementelor necesare analizei dispersionale.

Scanned with OKEN Scanner

Tabelul 5.11

Rezistența la forfecare (perpendicular pe fibre) la lemnul de molid, în raport cu tipul de pădure și poziția arborelui în arboret

Tipul de pădure (factorul *A*)	Poziția arborelui în arboret (factorul *B*)									
	1		2		3		4		5	
1	296	335	328	325	319	329	322	291	322	322
	323	327	316	339	285	322	286	326	317	322
2	307	306	333	328	320	287	321	306	316	311
	306	287	338	335	293	294	323	314	314	313
3	291	302	314	302	308	299	300	297	295	392
	296	290	291	302	312	304	293	302	310	318
4	284	298	295	301	286	287	302	312	281	286
	299	296	293	295	290	289	310	295	302	295
5	298	285	288	290	278	292	317	280	286	274
	281	296	300	295	286	295	314	312	293	290
6	280	301	306	293	299	271	300	295	297	282
	296	298	292	292	301	302	302	285	287	297

Tabelul 5.12

Datele din tabelul 5.11 micșorate cu 300 unități

Variantele factorului *A*	Variantele factorului *B*									
	1		2		3		4		5	
1	— 4	35	28	25	19	29	13	— 9	22	22
	23	27	16	39	—15	22	—14	26	17	22
2	7	6	33	28	20	—13	21	6	16	11
	6	— 3	33	35	— 7	— 6	23	14	14	13
3	— 9	2	14	2	8	— 1	0	— 3	— 5	— 8
	— 4	—10	— 9	2	12	4	— 7	2	10	18
4	—16	— 2	— 5	1	—14	—13	2	12	—19	—14
	— 1	— 4	— 7	— 5	—10	—11	10	— 5	2	— 5
5	— 2	—15	—12	—10	—22	— 8	17	—20	—14	—26
	—19	— 4	0	— 5	—14	— 5	14	12	— 7	—10
6	—20	1	6	— 7	— 1	—29	0	— 5	— 3	—18
	— 4	— 2	— 8	— 8	1	2	2	—15	—13	— 3

Numărul total al observațiilor: 120.

$$t = 6;\ b = 5;\ n = 4 \text{ repetiții}$$

$$N = 4\ tb = 4.6.5 = 120$$

$$G = 240$$

$$C = \frac{G^2}{N} = \frac{240^2}{120} = 480$$

Scanned with OKEN Scanner

$$Q = \Sigma x_{ij}^2 - C = [(-4)^2 + 35^2 + \ldots + (-3)^2] - 480 = 25\,942$$

$$Q_A = \Sigma \frac{T_i^2}{nb} - C = \frac{1}{20}(117\,649 + 66\,049 + \ldots + 15\,376) - 480 = 11\,156$$

$$Q_B = \Sigma \frac{B_j^2}{n \cdot t} - C = \frac{1}{24}(144 + 34\,596 + \ldots + 484) - 480 = 1\,484.$$

Suma pătratelor în raport cu mediile calculate pentru fiecare combinație se determină astfel:

$$Q_m = n\Sigma\bar{x}_{ij}^2 - C \qquad (5.14)$$

unde:

n este numărul repetițiilor pentru fiecare combinație;

$\bar{x}_{ij}$ — mediile pentru combinația ij.

$$Q_m = 4\left[\left(\frac{81}{4}\right)^2 + \left(\frac{108}{4}\right)^2 + \ldots \left(\frac{-37}{4}\right)^2\right] - 480 = 17\,088.$$

Tabelul 5.13

Datele din tabelul 5.12 însumate

Variantele factorului A	Variantele factorului B					Suma T_i
	1	2	3	4	5	
1	81	108	55	16	83	343
2	16	129	— 6	64	54	257
3	—21	9	23	— 8	15	18
4	—23	—16	—48	19	—36	—104
5	—40	—27	—49	23	—57	—150
6	—25	—17	—27	—18	—37	—124
Suma B_i	—12	186	—52	96	22	240

Tabelul 5.14

Datele din tabelul 5.13 ridicate la pătrat

Variantele factorului A	Variantele factorului B					T_i^2
	1	2	3	4	5	
1	6 561	11 664	3 025	256	6 889	117 649
2	256	16 641	36	4 096	2 916	66 049
3	441	81	529	64	225	324
4	529	256	2 304	361	1 296	10 816
5	1 600	729	2 401	529	3 299	22 500
6	625	289	729	324	1 369	15 376
B_i^2	144	34 596	2 704	9 216	484	57 600

Scanned with OKEN Scanner

Suma pătratelor pentru *interacțiunea factorilor A* și *B* va fi (vezi tab. 5.15)

$$Q_{AB} = 17\,088 - (11\,156 + 1\,484) = 4\,448$$

iar

$$Q_R = 25\,942 - (4\,448 + 11\,156 + 1\,484) = 8\,854.$$

Schema analizei dispersionale este prezentată în tabelul 5.16, după modelul din tabelul 5.15.

Tabelul 5,15

Schema analizei dispersionale în cazul a doi factori cu *n* repetiții pentru fiecare combinație

Sursa împrăștierii	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul *F*
Factorul *A*	$Q_A = \Sigma \frac{T_i^2}{n\,b} - C$	$t-1$	$s_A^2 = \frac{Q_A}{t-1}$	$\frac{s_A^2}{s_R^2}$
Factorul *B*	$Q_B = \Sigma \frac{B_j^2}{n\,t} - C$	$b-1$	$s_B^2 = \frac{Q_B}{b-1}$	$\frac{s_B^2}{s_R^2}$
Interacțiunea *AB*	$Q_{AB} = Q_m - (Q_A + Q_B)$	$(t-1)\,(b-1)$	$s_{AB}^2 = \frac{Q_{AB}}{(t-1)\,(b-1)}$	$\frac{s_{AB}^2}{s_R^2}$
Reziduală	$Q_R = Q - (Q_{AB} + Q_A + Q_B)$	$N - tb$	$s_R^2 = \frac{Q_R}{N-tb}$	—
Totală	$Q = \Sigma x_{ij}^2 - C$	$N-1$	—	—

Tabelul 5.16

Analiza dispersională pentru datele din tabelul 5.11

Sursa împrăștierii	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul *F*
Tipul de pădure (factorul *A*)	11 156	5	2 231	22,7
Poziția arborelui (factorul *B*)	1 484	4	371	3,8
Interacțiunea *AB*	4 448	20	222	2,3
Reziduală	8 854	90	98	—
Totală	25 942	119	—	—

Scanned with OKEN Scanner

După tabelul din anexa 5 stabilim următoarele valori tabelare pentru criteriul F:

Factorul	f_1	f_2	Nivelul de semnificație, %	
			5	1
A	5	90	2,3	3,2
B	4	90	2,5	3,5
Interacțiunea AB	20	90	1,7	2,1

Pentru nivelul de semnificație de 5%, și chiar pentru cel de 1%, valorile tabelare ale criteriului F sînt mai mici decît valorile experimentale corespunzătoare, așa încît conchidem că toate sursele de variații (tipul de pădure, poziția arborelui în arboret și *interacțiunea acestor doi factori*) exercită o influență asupra rezistenței la forfecare (perpendicular pe fibre) la lemnul de molid experimentat.

În continuare se stabilește gradul de semnificație al diferențelor folosind testul t. Să stabilim, de exemplu, semnificația diferenței dintre mediile primelor două tipuri de pădure. Calculăm mai întîi mediile:

$$\bar{x}_1 = \frac{343}{20} = 17{,}15; \quad \bar{x}_2 = \frac{257}{20} = 12{,}85; \quad \bar{x}_1 - \bar{x}_2 = 4{,}30$$

Eroarea acestei diferențe este dată de formula:

$$s_d = s_R \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

unde: $s_R = \sqrt{s_R^2} = \sqrt{98} = 9{,}9.$

Prin urmare

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_R \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} = \frac{17{,}15 - 12{,}85}{3{,}1} = \frac{4{,}3}{3{,}1} = 1{,}4$$

Diferența este deci nesemnificativă (vezi anexa 7).

O analiză de ansamblu asupra semnificației diferențelor se poate efectua prin metoda DUNCAN (§ 5.2).

În experiențele prezentate, pentru fiecare combinație a factorului A cu factorul B s-au efectuat același număr de măsurători. În cazul cînd această condiție nu este îndeplinită, trebuie folosite procedee de calcul speciale (S n e d e c o r, 1957; L e o n t i e v, 1961 ș.a.).

Scanned with OKEN Scanner

5.4. ANALIZA DISPERSIONALĂ PENTRU EXPERIENȚE POLIFACTORIALE*

În tehnica experimentală modernă se introduc din ce în ce mai mult experiențele prin care se cercetează concomitent acțiunea mai multor factori. În comparație cu metodele clasice, experiențele polifactoriale prezintă multe avantaje, dintre care mai importante sînt următoarele:

1. Se obține imaginea influenței fiecărui factor în condiții diferite, create de schimbarea altor factori.

2. Folosirea diferitelor combinații dintre factori permite să se obțină o bază mai sigură pentru recomandări practice, recomandări care rămîn valabile și în cazul schimbării anumitor condiții.

3. În cazul cînd toți factorii acționează independent unul de altul, atunci prin această experiență obținem tot atîtea informații cîte am fi obținut în situația unei experiențe clasice.

4. Dacă factorii luați în considerare acționează în interdependență, atunci printr-o astfel de experiență se obțin informații suplimentare deosebit de utile privind *interacțiunea acestor factori.*

În teoria și practica organizării lucrărilor experimentale se cunosc multe scheme de experiențe polifactoriale, corespunzătoare anumitor scopuri de cercetare. În cele ce urmează ne vom opri numai la anumite idei de bază ale experiențelor polifactoriale, dînd un exemplu simplu dar frecvent întîlnit în practică. Cititorul interesat va găsi suficiente explicații în plus în lucrările citate la bibliografie (S ă u l e s c u, 1959; V e s s e r e a u, 1960; S n e d e c o r, 1957; W i s h a r t și S a n d e r s, 1955; H a l d, 1952 ș.a.).

Experiențe polifactoriale de tipul 2^n. Analizăm schema experiențelor cu n factori, unde fiecare factor prezintă două graduări. Prin graduare înțelegem anumite nivele cantitative, sau anumite concentrații (în cazul aplicării îngrășămintelor). Uneori prin graduări se pot înțelege obiecte calitativ diferite, de exemplu două specii diferite. Prin graduări în anumite situații se înțelege prezența sau absența factorului respectiv.

De obicei, factorii se notează prin A,B,C ș.a., iar prin a, b, c ș.a. — se notează *una din cele două* graduări ale factorului corespunzător. Spre exemplu, dacă într-o plantație de plopi euramericani, în mod experimental, în anumite parcele acționăm prin trei factori (A — prelucrarea adîncă a solului, B — aplicarea de îngrășăminte și C —irigări) prin a vom înțelege acțiunea factorului A, prin b — acțiunea factorului B și prin c — acțiunea factorului C. Prin ab vom înțelege acțiunea comună a factorilor A și B, iar prin abc — acțiunea simultană a celor trei factori luați în considerare. Simbolul ac arată că în parcelele experimentale respective s-a aplicat prelucrarea adîncă a solului și în plus s-au efectuat irigări. Lipsa celor trei factori se notează prin „1". În experiența noastră „1" înseamnă varianta martor cu prelucrarea obișnuită a solului.

În cazul a trei factori este ușor de înțeles că vom avea 8 acțiuni, și anume: „1", a, b, ab, c, ac, bc, abc.

---

* După expunerea sistematică prezentată de B a i l e y (1959).

Scanned with OKEN Scanner

Desigur, fiecare caz poate avea mai multe repetiţii, aşezate în blocuri experimentale de diferite tipuri (§ 5.5). Să presupunem că am instalat r blocuri. Creşterea totală (dacă acesta este indicatorul urmărit în experimentarea culturilor comparative cu plopi euramericani) pentru cele r parcele unde acţionează „1", a, b, ab, c ş.a. o vom nota prin T_1, T_a, T_b, T_{ab}, T_c ş.a. Mediile obţinute prin împărţirea acestor sume prin r, se notează astfel: „$\bar{1}$", $\bar{a}$, $\bar{b}$, $\bar{ab}$, $\bar{c}$ ş.a.

Pe baza următorului exemplu vom explica modul de desfăşurare a analizei dispersionale în cazul experienţei polifactoriale de tipul 2^3 (trei factori cu 2 graduări fiecare).

În experienţa privind influenţa celor trei factori (A — prelucrarea adîncă a solului, B — aplicarea de îngrăşăminte şi C — irigare), asupra dezvoltării culturilor de plopi euramericani, s-au instalat 4 blocuri, fiecare bloc conţinînd 8 parcele (vezi § 5.5) .Variantele „1", a, b, ab, c, ac, bc, şi abc în cadrul fiecărui bloc au fost amplasate la întîmplare, folosind numerele întîmplătoare sau tragerea de bilete. După n ani de la instalarea culturilor au rezultat, pe parcele, producţiile totale (volumul arboretului principal plus volumul arboretului secundar) prezentate în tabelul 5.17.

Tabelul 5.17

Schema unei experienţe polifactoriale de tipul 2^3 cu 4 blocuri

Blocul 1

ab 280	„1" 100	*ac* 370	*a* 260
b 110	*abc* 450	*c* 300	*bc* 360

Blocul 2

ac 430	*abc* 430	*c* 320	„1" 105
a 280	*b* 95	*bc* 330	*ab* 310

Blocul 3

c 300	*ab* 320	„1" 80	*b* 130
bc 330	*a* 270	*ac* 400	*abc* 460

Blocul 4

bc 350	*ac* 420	*a* 300	*abc* 450
b 100	„1" 130	*ab* 275	*c* 310

Se pune acum în mod firesc întrebarea care sînt *efectele de bază şi efectele acţiunilor combinate* ale celor trei factori luaţi în considerare. O analiză simplă a rezultatelor obţinute nu poate rezista unei critici serioase din cauza neuniformităţii condiţiilor staţionale în care s-au instalat cele 4 blocuri experimentale. De aceea trebuie să apelăm la avantajele analizei dispersionale, care permite eliminarea variaţiei între blocuri.

În tabelul 5.18 se prezintă sumele producţiilor realizate pe blocuri şi pe variante.

Suma pentru varianta „1" s-a obţinut astfel:

$$100 + 105 + 80 + 130 = 415$$

Scanned with OKEN Scanner

Tabelul 5.18

Suma producțiilor realizate pe blocuri și pe variante

Acțiuni	Blocul				Suma
	1	2	3	4	
„1“	100	105	80	130	415
a	260	280	270	300	1 110
b	110	95	130	100	435
ab	280	310	320	275	1 185
c	300	320	300	310	1 230
ac	370	430	400	420	1 620
bc	360	330	330	350	1 370
abc	450	430	460	450	1 790
Suma	2 230	2 300	2 290	2 335	9 155

Analiza dispersională se efectuează în două etape. În prima etapă se stabilește dacă sistemul este sau nu omogen, folosind procedeul clasic descris la § 5.3. În cazul cînd prin analiza dispersională astfel efectuată se ajunge la un F experimental mai mare decît F tabelar, se trece la o analiză de detaliu a sistemului experimental.

În cadrul primei etape stabilim următoarele elemente ale analizei dispersionale:

Numărul total al observațiilor: $8 \times 4 = 32$

Suma totală $G = 9155$

Factorul de corecție C

$$C = \frac{G^2}{N} = \frac{9155^2}{32} = 2\,619\,188$$

Suma totală a pătratelor

$$Q = (100^2 + 105^2 + 80^2 + \ldots + 430^2 + 460^2 + 450^2) - 2\,619\,188 = 442\,687$$

Suma pătratelor pentru blocuri

$$Q_B = \frac{1}{8}(2\,230^2 + 2\,300^2 + 2\,290^2 + 2\,335^2) - 2\,619\,188 = 715\ *.$$

Suma pătratelor între acțiuni

$$Q_A = \frac{1}{4}(415^2 + 1\,110^2 + \ldots + 1\,790^2) - 2\,619\,188 = 434\,831\ **.$$

Rezultatele obținute le trecem în tabelul analizei dispersionale (tabelul 5.19).

* S-a luat $t = 8$ deoarece avem 8 acțiuni.
** S-a luat $b = 4$, deoarece schema cuprinde 4 blocuri.

Scanned with OKEN Scanner

Tabelul 5.19

Analiza dispersională după datele tabelului *5.18*

Împrăștierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul *F*
Între blocuri	715	3	238	0,7
Între variante	434 831	7	62 119	183
Reziduală	7 141	21	340	—
Totală	442 687	31	—	—

Neomogenitatea sistemului sub raportul factorilor experimentați este dovedită, deoarece *F* experimental este cu mult mai mare decît *F* tabelar luat la 7 și 21 grade de libertate. Conchidem deci că factorii luați în considerare influențează asupra creșterii arboretelor experimentate. În schimb, între blocuri se constată o fluctuație redusă deoarece *F* calculat este inferior celui tabelar, luat la 3 și 21 grade de libertate (*F* tabelar este egal cu 3,1 la $q = 5\%$). Neomogenitatea condițiilor staționale nu are deci nici o influență asupra creșterii arboretelor studiate. Formarea blocurilor nu duce în acest caz la o îmbunătățire a rezultatelor.

Următoarea etapă trebuie să scoată în evidență efectul factorilor izolați și al acțiunii combinate a acestora.

Aceste calcule sînt destul de complicate. Pentru experiența de tipul 2_n s-au elaborat însă reguli clare care aplicate corect pot duce la rezultate bune într-un timp relativ scurt.

Calculele se efectuează după schema prezentată în tabelul 5.20.

Tabelul 5.20

Rezultatele calculelor pentru determinarea efectelor de bază și a efectelor acțiunilor comune

Acțiuni	Suma (din tab. 5.18)	I	II	III
„1“	415	+1 525	+3 145	+9 155
a	1 110	+1 620	+6 010	+2 255
b	435	+2 850	+1 445	+ 405
ab	1 185	+3 160	+ 810	+ 85
c	1 230	+ 695	+ 95	+2 865
ac	1 620	+ 750	+ 310	— 635
bc	1 370	+ 340	+ 55	+ 215
abc	1 790	+ 420	+ 30	— 25

În prima coloană se trec acțiunile și combinațiile respective în ordinea lor normală, în aceeași ordine ca și în tabelul 5.18. În coloana a doua se înregistrează sumele preluate din același tabel.

Pentru completarea coloanei I s-a stabilit următoarea regulă:

— primele patru valori sînt egale cu sumele perechilor succesive ale cifrelor din coloana anterioară: (415 + 1 110 = 1 525; 435 + 1 185 = = 1 620; 1 230 + 1 620 = 2 850; 1 370 + 1 790 = 3 160);

Scanned with OKEN Scanner

— ultimele patru valori ale coloanei I reprezintă diferența dintre aceleași perechi de cifre, întotdeauna cifra de sus scăzîndu-se din cea de jos (1 110 — 415 = 695; 1 185 — 435 = 750; 1 620 — 1 230 = 390; 1 790 — 1 370 = 420).

Coloana II se completează după aceeași regulă, însă de data aceasta se operează cu cifrele din coloana I, iar pentru coloana III se folosesc cifrele din coloana II.

Dacă se urmărește cunoașterea valorii reale a acestor efecte, atunci toate sumele calculate mai sus se împart la mărimea $r2^{n-1}$, unde r reprezintă numărul blocurilor, iar n - numărul factorilor. În exemplul considerat această mărime este egală cu $4 \times 2^2 = 16$. Astfel, efectul de bază A, adică creșterea medie realizată prin prelucrarea adîncă a solului minus creșterea medie pentru parcelele fără acțiunea factorului A este de $+2255 : 16 = +141$. Pentru fiecare efect, dispersia este dată de formula $s^2/r2^{n-2}$. Limitele intervalului de încredere se stabilesc după formulele cunoscute.

Calculele sînt efectuate corect dacă suma pătratelor din coloana III, începînd cu a împărțită la N este egală cu suma pătratelor între variante. Într-adevăr:

$$Q_A = \frac{1}{32}(2255^2 + 405^2 + \ldots + 215^2 - 25^2) = 434831.$$

Prima cifră obținută în coloana III reprezintă suma totală $G = 9\,155$; aceasta trebuie să corespundă cu suma totală din tabelul 5.18.

Celelalte cifre din aceeași coloană (III) reflectă efectele de bază și efectele acțiunilor comune ale celor trei factori. Acestea servesc pentru analiza dispersională de detaliu. În acest scop se determină dispersia pentrn toate sumele din coloana III, după formula:

$$Ns_R^2 \qquad (5.15)$$

unde:

N este numărul total al observațiilor:

s_R^2 — dispersia reziduală din tabelul 5.19.

Eroarea va fi:

$$\sqrt{N\,s_R^2} = \sqrt{32 \times 340} = 104$$

Tabelul 5.20a

Acțiuni	Efectul (sporul)	Eroarea	Testul t
a	2 255	104	21,7
b	405	104	3,9
ab	85	104	0,8
c	2 865	104	27,5
ac	—635	104	6,1
bc	215	104	2,1
abc	—25	104	0,2

Gradul de semnificație a efectelor obținute se determină cu ajutorul testului t în raport cu cifrele din coloana III. După datele din tabelul 5.20 obținem rezultatele din tabelul 5.20 a.

După tabelul din anexa 7, folosind 21 grade de libertate, stabilim că:

1. efectele de bază (al prelucrării adînci a solului, aplicării îngrășămintelor și al irigării) sînt foarte semnificative;

Scanned with OKEN Scanner

2. efectul negativ al acțiunii combinate *ac* (prelucrarea adîncă a solului și irigarea) de asemenea este foarte semnificativ;

3. efectul acțiunii combinate al factorilor *B* și *C* (aplicarea îngrășămintelor și irigarea) este semnificativ;

4. efectele acțiunilor comune *ab* și *abc* nu sînt semnificative.

O condiție importantă pentru reușita experiențelor factoriale este omogenitatea condițiilor în care are loc experiența. Pentru experiențele din cîmp, acestea trebuie așezate pe terenuri cît se poate de omogene.

La experiențele unde măcar unul din factori prezintă mai mult de două trepte, metoda descrisă nu mai este aplicabilă. În acest caz se folosesc metode adecvate. Unele dintre acestea sînt prezentate de S ă u l e s c u (1959) după M u d r a (1952). Tot în această lucrare se prezintă și modalitatea de valorificare a experiențelor în serie (experiențe repetate în diferite localități și diferiți ani etc). Utilă este și lucrarea lui V e s s e r e a u (1960).

*

Uneori, în practica analizei dispersionale, pentru a evita anumite inconveniente legate de neomogenitatea dispersiilor sau pentru a corecta lipsa aditivității, se folosesc anumite transformări ale valorilor inițiale. Astfel, cînd se lucrează cu procente (x) se folosește transformarea $y = arc\ sin\ \sqrt{x}$ prezentată în tabele speciale (S n e d e c o r, 1957; E. W e b e r, 1957). Alteori este utilă transformarea $y = \sqrt{x}$ sau $y = \sqrt{0{,}5 + x}$, ori utilizarea logaritmilor.

Testul T u c k e y ne arată dacă transformarea este eficace și ne indică ce metodă de transformare să aplicăm (S n e d e c o r, 1957)

5.5. AȘEZAREA EXPERIENȚELOR

Pentru cercetările agricole s-au elaborat o serie de metode adecvate de așezare a experiențelor. Acestea au menirea să evite fluctuațiile datorite factorilor secundari neluați în seamă. Deși inițial au fost elaborate pentru experiențele agricole, pot fi cu succes adaptate și în alte domenii. Deosebit de utile sînt la instalarea experiențelor cu culturi forestiere comparative, la evidențierea efectului îngrășămintelor asupra dezvoltării culturilor forestiere, la stabilirea celor mai eficace metode de combatere a dăunătorilor, în experiențele privind regenerarea arboretelor, la amplasarea parcelelor experimentale privind aplicarea tăierilor de îngrijire, pentru determinarea preciziei diferitelor metode de cubaj, în experiențele privind efectul ierbicidelor în pepiniere și plantații etc. Multe experiențe de laborator, de asemenea, trebuie planificate cu luarea în considerare a fluctuațiilor generate de factorii secundari. Amintim aici experiențele privind studiul caracteristicilor fizico-mecanice ale lemnului, încercarea anumitor mașini și utilaje, studiul calității unor produse industriale etc.

Spre deosebire de metoda de lucru din agricultură, în experimentările din domeniul silviculturii și în deosebi în cele privind tăierile de îngrijire a arboretelor (E r t e l d, 1958), studiul creșterilor, aplicarea tratamen-

telor, folosirea îngrășămintelor (T h o m a s i u s, 1964), sau în cele privind culturile comparative, se întîmpină serioase dificultăți la formarea unor blocuri experimentale cu mai multe repetiții. Aceste dificultăți sînt generate de neuniformitatea condițiilor staționale pe suprafețe relativ mari, astfel încît fiecărei parcele experimentale să i se asigure o suprafață care la exploatabilitatea arboretului să cuprindă cel puțin 80—100 arbori. În cele mai multe experimentări instalate pînă în prezent un anumit tratament a fost prevăzut o singură dată, ceea ce — desigur — constituie un însemnat neajuns cînd se încearcă asigurarea rezultatelor prin metode moderne ale statisticii-matematice.

Cu toate aceste dificultăți, considerente de natură statistică obligă la formarea unor blocuri experimentale unde variantele să fie prevăzute în mai multe repetiții. După A s s m a n n (1961) o soluție ar fi dacă în experimentările noi pietele de experiență s-ar împărți în două ori în patru suprafețe parțiale, care s-ar inventaria și s-ar calcula separat atîta vreme cît ar permite numărul de arbori din aceste piețe. În acest mod devine posibilă interpretarea și asigurarea statistică a rezultatelor cel puțin pentru primele decenii de cercetare.

Dintre cele mai folosite metode de așezare a experiențelor menționăm următoarele: metoda blocurilor, pătratul latin, dreptunghiul latin, metoda grilajelor.

La alegerea celei mai potrivite metode sînt utile următoarele recomandări:

— la variante puține este indicată metoda blocurilor sau pătratul latin;
— la un număr mare de variante este mai potrivită metoda grilajelor;
— la soluri neuniforme sînt de preferat metodele: pătratul latin, dreptunghiul latin, precum și metoda grilajelor;
— în cazul unei forme alungite a terenului este de preferat așezarea liniară a repetițiilor.

Metoda blocurilor. Pentru experiențele agricole blocul reprezintă o grupare de parcele experimentale de mărime egală, așezate pe un sol cît mai uniform din punctul de vedere al fertilității. Se pot forma mai multe asemenea blocuri.

După așezare, blocurile pot fi:

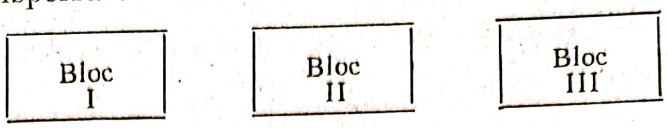
a) pe un singur rînd (așezare liniară)

Bloc I	Bloc II	Bloc III	Bloc IV	Bloc V

b) blocuri etajate

Bloc I	Bloc II	Bloc III
Bloc IV	Bloc V	Bloc VI

c) blocuri dispersate



Fiecare bloc conține atîtea parcele cîte variante se experimentează, cu condiția ca în interiorul fiecărui bloc variantele să fie repartizate *la întîmplare* (așezare *randomizată*). În acest scop se pot folosi două procedee: a) prin tragere la sorți; b) cu ajutorul tabelului cu numere întîmplătoare (anexa 9).

Numărul repetițiilor poate varia în funcție de necesitățile impuse, numărul de variante nu trebuie însă să fie prea mare (>15). Metoda asigură rezultate concludente atunci cînd terenul pe care s-au instalat experiențele este relativ uniform din punctul de vedere al fertilității solului.

Valorificarea experiențelor așezate în blocuri se efectuează după schema prezentată la § 5.3.

În tabelul 5.21 se prezintă rezultatele unei experiențe cu 6 blocuri și 4 variante, privind creșterea în înălțime la puieți de proveniență diferită.

Tabelul 5.21

Blocul	Înălțimile puieților pe variantele *A*, *B*, *C* și *D*			
1	B 40	D 38	C 40	A 39
2	D 43	A 42	C 42	B 43
3	A 42	C 42	B 42	D 43
4	C 42	B 40	D 41	A 43
5	B 38	A 38	D 37	C 38
6	A 38	C 36	D 40	B 42

În tabelul 5.22 valorile obținute se înscriu pe blocuri și variante.

Tabelul 5.22

Așezarea datelor din tabelul *5.21* pe blocuri și variante

Varianta	Blocul						Suma
	1	2	3	4	5	6	
A	39	42	42	43	38	38	242
B	40	43	42	40	38	42	245
C	40	42	42	42	37	36	239
D	38	43	43	41	38	40	243
Suma	157	170	169	166	151	156	969

Urmînd indicațiile de la § 5.3 rezultă datele din tabelul 5.23.

Conchidem deci că provenienta puieților nu influențează asupra creșterii în înălțime. Semnificative sînt însă diferențele între blocuri, din cauza neuniformității condițiilor staționale (F tabelar, pentru $q=5\%$, la $f_1=5$ și $f_2=15$ grade de libertate, reprezintă 2,9, fiind mai mic decît F calculat). De aici rezultă eficiența formării blocurilor.

Scanned with OKEN Scanner

Tabelul 5.23

Rezultatele analizei dispersionale după datele din tabelul 5.22

Împrăştierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Criteriul *F*
Intre variante (proveniențe)	4	3	1,3	0,6
Intre blocuri	79	5	15,8	7,2
Reziduală	33	15	2,2	—
Totală	116	23	—	—

Pătratul latin. Potrivit acestei metode, fiecare coloană şi fiecare rînd conține toate variantele luate în considerare. Deci fiecare variantă este prezentă o singură dată atît în fiecare bloc cît şi în fiecare coloană. Randomizarea se realizează nu pentru parcele, ci pentru blocurile şi coloanele planului de bază. Numărul variantelor este egal cu numărul repetițiilor.

Avantajul principal al acestei metode constă în aceea că permite un control temeinic al neuniformității condițiilor în care se realizează experimentarea. Din acest punct de vedere, verificarea se face atît pe coloane cît şi pe blocuri.

Una din multiplele modalități de aşezare a parcelelor după metoda pătratului latin se prezintă astfel (5 variante şi 5 blocuri)*:

blocul 1	*A*	*B*	*C*	*D*	*E*
blocul 2	*E*	*A*	*B*	*C*	*D*
blocul 3	*D*	*E*	*A*	*B*	*C*
blocul 4	*C*	*D*	*E*	*A*	*B*
blocul 5	*B*	*C*	*D*	*E*	*A*

Randomizarea se realizează prin metode cunoscute.

Metoda prezentată este potrivită pentru un număr relativ mic de variante (<12). Este frecvent folosită la planificarea experiențelor în serie. În silvicultură poate fi cu succes aplicată în multe experiențe de cîmp şi de laborator.

E x e m p l u. Să presupunem că se urmăreşte cunoaşterea influenței tipului de pădure şi a vîrstei arborilor asupra greutății semințelor de pin negru. În acest scop dispunem de material provenit din 4 regiuni. Organizăm efectuarea măsurătorilor în schema pătratului latin (tabelul 5.24), astfel încît atît pe coloane, cît şi pe rînduri fiecare tip de pădure să fie reprezentat cîte o singură dată.

* S ă u l e s c u (1959) prezintă scheme tip pentru experiențele aşezate în pătratul latin.

Scanned with OKEN Scanner

Tabelul 5.24

Greutatea semințelor de pin negru în raport cu vîrsta, tipul de pădure și proveniența (unitatea statistică = greutatea a *1 000* de semințe) Așezare după schema pătratului latin

Proveniența	Vîrsta arborilor			
	V_1	V_2	V_3	V_4
1	*A* 20	*B* 23	*C* 19	*D* 20
2	*B* 19	*C* 20	*D* 18	*A* 17
3	*C* 21	*D* 24	*A* 19	*B* 21
4	*D* 20	*A* 20	*B* 20	*C* 22

Cei doi factori (tipul de pădure și vîrsta arborilor) au cîte 4 nivele. Cele 4 tipuri de pădure sînt însemnate prin simbolurile *A*, *B*, *C* și *D*.

Sumele pe tipuri de pădure, vîrste și provenienţe sînt prezentate în tabelul 5.25.

Tabelul 5.25

Sumele pe rînduri, coloane și variante (provenienţe), după datele din tabelul *5.24*

Sumele pe	1	2	3	4
rînduri (provenienţe)	82	74	85	82
coloane (vîrste)	80	87	76	80
variante (tipuri de pădure)	76	83	82	82
Suma totală	323		Media	20,2

Cu ajutorul datelor din acest tabel calculăm elementele necesare analizei dispersionale (§ 5.3).

$$N = bt = 4 \times 4 = 16; \quad G = 323; \quad C = \frac{323^2}{16} = 6\,520$$

$$Q_{total} = (20^2 + 23^2 + 19^2 + 20^2 + 19^2 + 20^2 + \ldots + \ldots + 20^2 + 22^2) - 6\,520 = 47$$

$$Q_{rînduri} = \frac{1}{4}(82^2 + 74^2 + 85^2 + 82^2) - 6\,520 = 17$$

$$Q_{coloane} = \frac{1}{4}(80^2 + 87^2 + 76^2 + 80^2) - 6\,520 = 16$$

$$Q_{variante} = \frac{1}{4}(76^2 + 83^2 + 82^2 + 82^2) - 6\,520 = 8$$

Suma pătratelor pentru dispersia reziduală se obţine prin diferenţă:

$$Q_R = 47 - (17 + 16 + 6) = 6$$

Cu aceste date se întocmește tabelul analizei dispersionale (tabelul 5.26).

Scanned with OKEN Scanner

Tabelul 5.26

Împrăștierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul F
Intre rînduri (proveniență)	17	3	5,7	5,7
Intre coloane (vîrste)	16	3	5,3	5,3
Intre variante (tipuri de pădure)	8	3	2,7	2,7
Reziduală	6	6	1,0	—
Totală	47	15	—	—

Gradele de libertate se calculează astfel:

— pentru dispersia totală $N-1=16-1=15$

— pentru dispersia între rînduri $4-1=3$

— pentru dispersia între coloane $4-1=3$

— pentru dispersia între varianta $4-1=3$.

Pentru dispersia reziduală numărul gradelor de libertate rezultă scăzînd suma gradelor de libertate calculate pentru rînduri, coloane și variante din numărul gradelor de libertate al dispersiei totale: $15-(3+3+3)=6$.

Analizînd rezultatele obținute se constată că cea mai mare parte din fluctuație se datorește provenienței, vîrstei și tipului de pădure. Dispersia reziduală este relativ mică, ceea ce arată eficacitatea planificării experimentării în acest mod. Valoarea F tabelară (luată la nivelul de semnificație de 5%, pentru $f_1=3$ și $f_2=6$) este de 4,8. De aici conchidem că numai între proveniențe și vîrste există diferențe care depășesc limitele întîmplării. Între tipurile de pădure diferențele sînt nesemnificative (fiindcă F tabelar este mai mare decît F calculat).

Desigur, pentru a întări concluzia privind lipsa unor diferențe semnificative pe tipuri de pădure, se impune repetarea unor astfel de experiențe.

Gradul de semnificație al diferențelor se calculează după procedee cunoscute, folosind testul t. Eroarea diferențelor va fi:

$$s_d = \sqrt{\frac{2s_R^2}{n}} = \sqrt{\frac{2\times 1,0}{4}} = 0,71$$

unde n reprezintă numărul variantelor.

Eroarea medie este

$$\sqrt{\frac{s_R^2}{n}} = \sqrt{\frac{1,0}{4}} = 0,50, \text{ iar } t = \frac{x_i - x_j}{s_d}$$

O analiză de ansamblu obținem prin metoda DUNCAN (§ 5.2).

Dreptunghiul latin. Spre deosebire de pătratul latin, blocurile nu formează coloane întregi care să cuprindă toate variantele așezate pe un singur șir vertical. O astfel de așezare permite includerea unui număr mai mare de variante.

Scanned with OKEN Scanner

La așezarea în dreptunghi latin se impun următoarele condiții:

— numărul coloanelor trebuie să fie egal cu numărul blocurilor (repetițiilor);

— numărul variantelor trebuie să fie divizibil cu numărul repetițiilor;

— fiecare variantă este reprezentată o singură dată în fiecare coloană și în fiecare bloc.

Numărul de variante nu trebuie să depășească 15—25. Este folosit atunci cînd pătratul latin nu poate fi aplicat din cauza unui număr mare de repetiții. Randomizarea se realizează separat pe blocuri și coloane. Dăm mai jos un exemplu de așezare în dreptunghi latin pentru 8 variante (4×2) cu 4 blocuri*.

Tabelul 5.27

Coloana	Bloc 1		Bloc 2		Bloc 3		Bloc 4	
1	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
2	*E*	*G*	*F*	*B*	*A*	*H*	*D*	*C*
3	*C*	*F*	*H*	*A*	*D*	*G*	*B*	*E*
4	*H*	*D*	*E*	*G*	*B*	*C*	*A*	*F*

Prelucrarea datelor experimentale pentru experiențele așezate în dreptunghi latin nu diferă față de valorificarea specifică pătratului latin. Numărul gradelor de libertate se calculează astfel:

— pentru dispersia totală: $N - 1$

— pentru dispersia între rînduri: $b - 1$

— pentru dispersia între coloane: $t - 1$

— pentru dispersia între variante: $v - 1$

— pentru dispersia reziduală — prin diferență.

Prin N s-a notat numărul total al parcelelor, b — reprezintă numărul blocurilor, t — numărul rîndurilor, iar v — reprezintă numărul variantelor.

Alte scheme de așezare a experiențelor. În cazul unor experiențe cu un număr mare de variante, schemele descrise mai sus devin incomode. De aceea s-au elaborat metode mult mai perfecte (grilajul balansat, grilajul parțial balansat, grilajul pătrat etc.). Valorificarea rezultatelor experiențelor instalate după aceste scheme este însă destul de complicată, de cele mai multe ori fiind necesară consultarea unui specialist în acest domeniu. Schemele sînt pe larg descrise în lucrările: Vessereau (1960), Wishart-Sanders (1955), Snedecor (1957), Mudra (1958), Weber (1957) ș.a. Pentru probleme de genetică sînt utile lucrările Kempthorne (1953), Wright Jonathan (1963) ș. a. Metodele grilajelor pătrate și balansate sînt prezentate pentru cercetările agricole de Săulescu (1959).

* Săulescu (1959) prezintă mai multe scheme tip pentru așezarea experiențelor în dreptunghi latin.

Scanned with OKEN Scanner

În principiu, sînt de preferat schemele mai simple cu un număr redus de variante, mai ales în cazul experienţelor polifactoriale. Foarte importantă este condiţia omogenităţii terenului în care sînt aşezate experienţele.

5.6. ANALIZA COVARIANŢE

Analiza dispersională nu este indicată în cazul în care caracteristica cercetată depinde de un alt factor, respectiv atunci cînd această caracteristică este corelată liniar cu un anumit factor bine determinat. În asemenea situaţii diferenţele dintre mediile colectivităţilor luate în studiu se pot datora nu factorilor urmăriţi prin cercetare, ci unor factori cu care caracteristica respectivă este corelată. Pentru înlăturarea acestor influenţe perturbante se foloseşte *analiza covarianţei.*

Spre deosebire de analiza dispersională, de data aceasta, se compară nu două dispersii faţă de mediile respective, ci două dispersii faţă de *ordonatele liniei de regresie* a lui y faţă de x*. Se demonstrează că suma pătratelor abaterilor faţă de ordonatele dreptei de regresie este dată de formula:

$$Q_y/\hat{y} = Q_y - \frac{Q_{xy}^2}{Q_x} \tag{5.16}$$

unde:

Q_y = suma pătratelor variabilei y;

Q_x = suma pătratelor variabilei x

$$Q_{xy} = \sum xy - \frac{\Sigma\, x\, \Sigma\, y}{n} \tag{5.17}$$

Analiza covarianţei se va ilustra în baza următorului exemplu:

Pentru un arboret echien de stejar se urmăreşte stabilirea influenţei tratamentelor aplicate asupra creşterii arborilor în diametru. În total s-au experimentat patru variante. Pentru fiecare variantă la arborii rămaşi s-au măsurat creşterile în diametru, realizate după 5 ani de la prima intervenţie. Totodată s-au înregistrat şi diametrele iniţiale ale coroanelor la arborii respectivi, deoarece se cunoaşte influenţa coroanei asupra creşterilor în diametru. Rezultatele obţinute sînt înregistrate în tabelul 5.28.

Efectuăm iniţial o analiză a împrăştierii fără a lua în considerare factorul coroană. În acest scop folosim indicaţiile date la § 5.2.

Pentru variabila y (creşterea în diametru) avem:

$$Q_{y\ (\text{total})} = (6^2 + 9^2 + 11^2 + \ldots + 7^2 + 8^2) - \frac{242^2}{33} = 71{,}333$$

$$Q_{y\ (\text{între variante})} = \frac{87^2}{10} + \frac{45^2}{7} + \frac{60^2}{9} + \frac{50^2}{7} - \frac{242^2}{33} = 28{,}662$$

* Vezi cap. 6 şi 7

Tabelul 5.28

Variația creșterilor în diametru la arbori în urma răririi arboretului prin tăieri de regenerare aplicate în patru variante

Varianta								Total	
A		*B*		*C*		*D*			
Creșterea în diametru, mm	Diametrul coroanei, m	Creșterea în diametru, mm	Diametrul coroanei, m	Creșterea în diametru, mm	Diametrul coroanei, m	Creșterea în diametru, mm	Diametrul coroanei, m	Creșterea în diametru, mm	Diametrul coroanei, m
6	4,0	6	4,5	7	4,5	6	6,0		
9	6,5	6	4,0	7	4,0	6	5,0		
11	6,0	7	5,5	5	4,5	7	6,5		
9	6,0	6	3,0	7	5,0	8	4,5		
10	5,0	7	3,5	6	2,5	8	4,0		
7	4,5	7	4,0	9	6,0	7	5,5		
10	5,5	6	3,0	7	2,5	8	4,5		
8	5,5			5	2,5				
10	6,0			7	6,0				
7	5,0								
87	54,0	45	26,5	60	37,5	50	38,0	242	156,0

Din tabelul analizei dispersionale rezultă (tabelul 5.29):

$$F = \frac{9{,}554}{1{,}472} = 6{,}5.$$

Tabelul 5.29

Analiza dispersională pentru variabila *y* (după datele din tabelul *5.28*)

Împrăștierea	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul *F*
Intre variante	28,662	3	9,554	6,5
Reziduală	42,671	29	1,472	
Totală	71,333	32		

Din tabelul anexei 5, la $f_1 = 3$ și $f_2 = 29$, și la nivelul de semnificație de 0,05, stabilim că valoarea tabelară a testului F este de 2,9. Conchidem deci, că din punct de vedere al creșterii în diametru există diferențe semnificative între cele 4 variante experimentale. Concluzia rămîne valabilă și pentru nivelul de semnificație de 0,01 (testul F tabelar este de 4,6).

Pentru o analiză corectă a influenței tratamentului aplicat asupra creșterii în diametru va trebui să eliminăm influența factorului „diametrul coroanei", adică să aducem toți arborii la același nivel din punct de vedere al diametrului inițial al coroanei. Acest lucru este necesar deoarece anumite diferențe între variantele aplicate se pot datora nu tratamentelor respective, ci faptului că există deosebiri inițiale în ceea ce privește dimen-

Scanned with OKEN Scanner

siunile coroanei. În acest scop folosim analiza covarianței, determinînd suma pătratelor abaterilor față de ordonatele dreptei de regresie (formula 5.16). Calculăm deci Q_x și Q_{xy}.

Pentru factorul „y" (diametrul coroanei) avem:

$$Q_{x\ (\text{total})} = (4{,}0^2 + 6{,}5^2 + \ldots + 4{,}5^2) - \frac{156^2}{33} = 48{,}546$$

$$Q_{x(\text{între variante})} = \frac{54^2}{10} + \ldots + \frac{38^2}{7} - \frac{156^2}{33} = 17{,}003$$

Așa încît pentru variabila x obținem:

Împrăștierea	Suma pătratelor
Intre variante	17,003
Reziduală	48,546 — 17,003 = 31,543
Totală	48,546

Pentru xy rezultă:

$$Q_{xy\ (\text{total})} = \Sigma xy - \frac{\Sigma x\, \Sigma y}{n} = (6\times 4{,}0 + 9\times 6{,}5 + 11\times 6{,}0 + \ldots + 8\times 4{,}5) - \frac{242\times 156}{33} = 33{,}000$$

$$Q_{xy\ (\text{între variante})} = \left(\frac{87\times 54{,}0}{10} + \ldots + \frac{50\times 38{,}0}{7}\right) - \frac{242\times 156}{33} = 17{,}586$$

Deci, pentru xy avem:

Împrăștierea	Suma pătratelor
Între variante	17,586
Reziduală	15,414
Totală	33,000

Cu rezultatele obținute întocmim următorul centralizator:

Tabelul 5.30

Împrăștierea	Q_x	Q_y	Q_{xy}	Numărul gradelor de libertate
Intre variante	17,003	28,662	17,586	3
Reziduală	31,543	42,671	15,414	29
Totală	48,546	71,333	33,000	32

Scanned with OKEN Scanner

Aplicînd formula (5.16) obţinem sumele pătratelor faţă de dreapta de regresie:

$$Q_{y/\hat{y}\ (\text{total})} = Q_{y\ (\text{total})} - \frac{Q^2_{xy\ (\text{total})}}{Q_{x(\text{total})}} = 71{,}333 - \frac{33{,}000^2}{48{,}546} = 48{,}901$$

$$Q_{y/\hat{y}\ (\text{reziduală})} = 42{,}671 - \frac{15{,}414^2}{31{,}542} = 35{,}138$$

În sfîrşit, întocmim tabelul final al analizei dispersionale (în raport cu dreapta de regresie):

Tabelul 5.31

Abaterea faţă de dreapta de regresie	Suma pătratelor	Numărul gradelor de libertate	Dispersia	Testul F
Totală	48,901	31		
Reziduală	35,138 (prin diferenţă)	28	1,25	3,67
Intre variante	13,763	3	4,59	

Numărul gradelor de libertate pentru dispersia totală este de $n - 2$ (în cazul nostru $33 - 2 = 31$). Între variante avem $f = 3$ (numărul variantelor minus unu).

La $f_1 = 3$ şi $f_2 = 28$ grade de libertate, pentru nivelul de semnificaţie de 0,05, stabilim că F tabelar este de 3,0. Deoarece testul F tabelar este mai mic decît testul F calculat, conchidem că între cele patru variante, la nivelul de semnificaţie de 5%, sînt diferenţe statistic asigurate, chiar şi după ce eliminăm influenţa datorată deosebirilor de coroană. Pentru nivelul de semnificaţie de 1%, testul F este însă de 4,6, fiind deci mai mare decît valoarea F experimentată. La acest nivel de semnificaţie diferenţele dintre variantele experimentate nu mai sînt statistic asigurate. Analiza dispersională efectuată anterior (tabelul 5.29) arată nejustificat că şi la acest nivel diferenţele sînt semnificative, fapt explicabil dacă avem în vedere că în acele calcule nu s-au luat în considerare dimensiunile iniţiale ale coroanelor.

*

Menţionăm că în practica lucrărilor statistice analiza covarianţei ocupă un loc de seamă avînd multiple aplicaţii. În paginile de faţă ne-am limitat la o prezentare simplificată, urmînd ca, la nevoie, cititorul să consulte lucrările citate la bibliografie (ex. Snedecor, 1957); Vessereau (1960). Exemple din silvicultură sînt prezentate de Tomassone (1964), Franz (1965) ş.a.

Metoda trebuie folosită la înlăturarea erorilor sistematice din cadrul blocurilor, la metoda blocurilor randomizate, în cazul folosirii pătratului latin, precum şi în multe alte diverse situaţii.

Scanned with OKEN Scanner

VI

DESPRE CORELAȚIE

6.1. GENERALITĂȚI

Analiza dispersională permite evidențierea influențelor anumitor factori asupra caracteristicii rezultative, însă nu determină gradul acestei influențe; nu spune nimic despre forma legăturii evidențiate. În scopul elucidării acestor aspecte cercetătorul va apela la *teoria corelației*.

Pentru corecta înțelegere a materialului ce se va prezenta în acest capitol, este necesar să arătăm deosebirile esențiale dintre *funcțiile matematice* și *funcțiile experimentale* (statistice). Corespondența $y = f(x)$, cum ar fi spre exemplu expresia $s = \pi r^2$, are caracterul unei funcții matematice, exacte, în sensul că unei anumite valori date lui x îi corespunde o valoare (sau mai multe valori dar bine determinate) a variabilei dependente. În cazul funcțiilor statistice legătura dintre varibila x și frecvențele respective nu mai are același caracter; pentru aceeași valoare a lui x putem avea valori diferite pentru y. De pildă, într-un arboret echien de molid, la același diametru d putem întîlni valori diferite pentru înălțimi, cu toate că pe măsură ce se mărește diametrul se majorează și media înălțimilor. În acest caz al funcțiilor experimentale (statistice), corespondența nu are loc de la valoare la valoare ci de la valoare la distribuție. Deci fiecărei valori x îi corespunde o anumită distribuție cu caracteristicile ei statistice (medie, dispersie etc.). Pe măsură ce se schimbă valoarea x se modifică și *centrul de grupare* (media) valorilor y; se modifică bineînțeles și dispersia corespunzătoare. Să analizăm tabelul 6.1 unde se prezintă legătura statistică dintre vîrsta și diametrul mediu la arboretele de molid din nordul țării, încadrate în tipul de pădure *Molidiș normal cu Oxalis acetosella*. Se vede că pe măsură ce se mărește x (vîrsta arboretelor) se modifică și distribuțiile pentru variabila y (diametrul mediu al arboretelor). Centrul de grupare al valorilor y își modifică treptat poziția; se modifică și amplitudinea împrăștierii valorilor y în jurul mediei $\overline{y}$.

Legăturile de acest gen se numesc *legături stochastice* sau *probabilistice*, încadrate în *teoria corelației*. Cu ajutorul teoriei corelației se pot scoate în evidență existența anumitor legități statistice. Dar asemenea legități apar în evidență numai în fenomenele de masă, făcînd observații numeroase. Studiul corelației în astfel de situații descoperă legăturile dintre fenomene,

Scanned with OKEN Scanner

pune în lumină forma acestor legături, arată gradul de intensitate al legăturii evidențiate, poate arăta ponderea diverșilor factori la realizarea unui fenomen de masă.

Tabelul 6.1

Tabel de corelație privind variația diametrelor medii în raport cu vîrsta arboretelor pentru tipul de pădure *Molidiș normal cu Oxalis acetosella* (după Giurgiu R. 1962)

Diametre medii $\bar{y}$ cm	Vîrsta ani, x													n
	5	15	25	35	45	55	65	75	85	95	105	115	135	
8	4	9	—	—	—	—	—	—	—	—	—	—	—	13
10	3	43	12	—	—	—	—	—	—	—	—	—	—	58
12	—	11	14	—	—	—	—	—	—	—	—	—	—	25
14	—	19	48	7	—	—	—	—	—	—	—	—	—	74
16	—	4	77	17	—	—	—	—	—	—	—	—	—	98
18	—	—	42	37	5	1	—	—	—	—	—	—	—	85
20	—	—	13	87	18	6	—	—	—	—	—	—	—	124
22	—	—	3	59	30	13	1	—	—	—	—	—	—	106
24	—	—	2	42	53	35	2	—	—	—	—	—	—	134
26	—	—	—	21	79	31	17	3	—	—	—	—	—	161
28	—	—	—	14	48	69	29	10	1	—	—	—	—	171
30	—	—	—	6	39	73	35	25	7	1	—	—	—	186
32	—	—	—	—	9	34	59	45	12	6	—	—	1	166
34	—	—	—	—	11	17	30	40	26	12	3	4	—	143
36	—	—	—	—	3	12	21	35	32	23	9	5	2	142
38	—	—	—	—	—	2	9	18	19	22	3	2	—	75
40	—	—	—	—	—	—	—	8	6	30	10	9	3	66
42	—	—	—	—	—	—	2	5	5	9	12	11	2	46
44	—	—	—	—	—	—	—	2	3	9	16	5	4	39
46	—	—	—	—	—	—	—	1	1	2	9	9	3	25
48	—	—	—	—	—	—	—	—	—	—	3	7	1	11
50	—	—	—	—	—	—	—	—	—	—	2	9	5	16
52	—	—	—	—	—	—	—	—	—	—	1	2	6	9
54	—	—	—	—	—	—	—	—	—	—	—	2	2	4
56	—	—	—	—	—	—	—	—	—	—	—	1	6	7
58	—	—	—	—	—	—	—	—	—	—	—	1	1	2
60	—	—	—	—	—	—	—	—	—	—	—	—	2	2
62	—	—	—	—	—	—	—	—	—	—	—	—	3	3
64	—	—	—	—	—	—	—	—	—	—	—	—	2	2
66	—	—	—	—	—	—	—	—	—	—	—	—	—	—
68	—	—	—	—	—	—	—	—	—	—	—	—	1	1
Total	7	86	211	290	295	303	204	192	113	114	68	67	44	1994
$\bar{x}$	8,8	11,2	15,8	21,4	26,2	28,2	31,5	34,0	35,7	38,3	42,0	44,2	50,0	—
s	2,2	2,8	3,0	3,5	3,7	3,7	3,4	3,5	3,9	3,3	4,2	5,5	8,3	—
$s\%$	25	25	26	17	15	13	11	10	10	9	10	12	16	

coeficientul de corelație $r = 0,89 \pm 0,04$

Legăturile stochastice se pot exprima grafic (prin curbe) sau analitic, prin *ecuații de regresie*. Aceste ecuații de regresie au însă alt caracter decît ecuațiile matematice. Cu ajutorul lor nu obținem valori individuale exacte, obținem în schimb centrele de grupare, respectiv mediile unei serii de distribuții ale caracteristicii considerate.

În tabelul 6.1 s-au înregistrat frecvențele care se referă la cele două caracteristici: vîrsta arboretelor și diametrul mediu. Distribuția obținută se referă deci la două caracteristici; de aici denumirea de *distribuție bidimensională* (sau de două variabile). Tabelar aceasta se prezintă sub forma unui *tabel de corelație*. Pe grafic, unde pe axa x trecem variabila independentă, iar pe axa y — variabila dependentă, obținem un *cîmp de corelație*.

Pentru fiecare valoare (clasă) a lui x (vîrsta arboretelor) avem o distribuție a lui y (diametrul mediu al arboretelor) și o medie corespunzătoare a lui y. Înlocuind distribuțiile prin valorile medii $\overline{y}$ obținem corespondența:

x_i	5	15	25	35	...
$\overline{y}_i$	8,8	11,2	15,8	21,4	....

Această legătură se poate reprezenta grafic; obținem astfel o linie poligonală denumită *linie de regresie a lui y în raport cu x* (fig. 6.1). Această linie reprezintă o aproximație a unei linii teoretice.

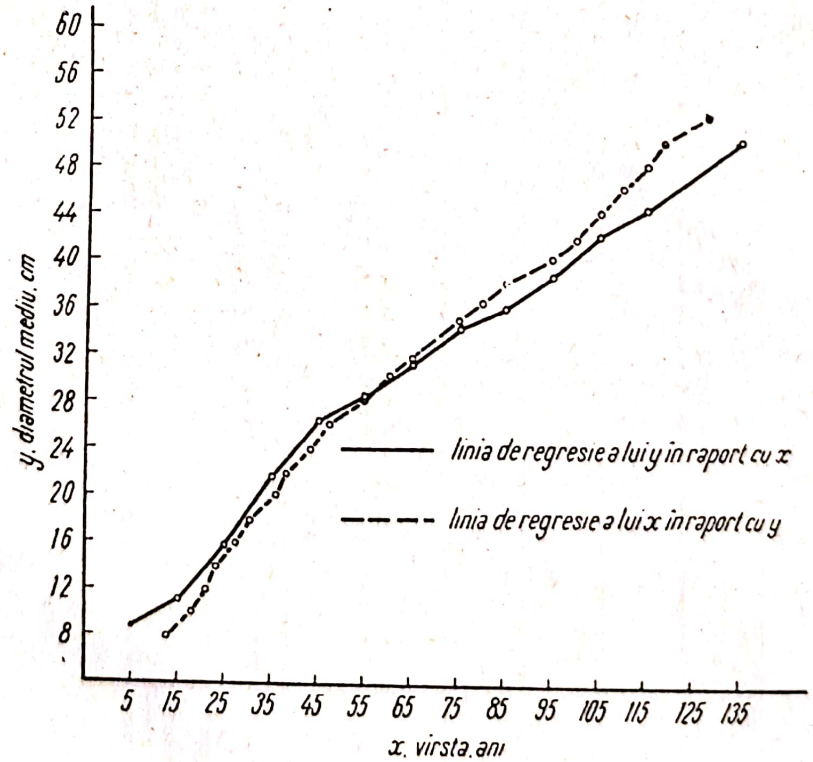


Fig. 6.1. Corelația dintre vîrsta (x) și diametrul mediu (y) la arboretele echiene de molid din tipul de pădure *Molidiș normal cu Oxalis acetosella*.

Scanned with OKEN Scanner

În mod analog calculăm mediile pe orizontale, adică mediile variabilei x (vîrsta arboretelor) pentru fiecare valoare a lui y (diametrul mediu):

y_i	8	10	12	14....
$\bar{x}_i$	11,9	16,6	20,6	23,3....

Pe baza acestor date obținem *linia de regresie a lui x în raport cu y*. Din grafic se vede clar că cele două linii nu se confundă ca în cazul ecuațiilor funcționale. Dacă legătura y față de x se exprimă analitic, prin intermediul unui polinom, în acest caz (al funcțiilor de regresie) nu putem calcula pe x față de y printr-o simplă operație de inversiune a funcțiilor. Aspectul subliniat trebuie mereu avut în vedere atunci cînd folosim curbe sau ecuații de regresie.

În cazul distribuțiilor bidimensionale ne găsim în situația unei *corelații simple*. Studiul corelației se complică o dată cu creșterea numărului de caracteristici luate în considerare (cazul corelației *multiple*).

În continuare ne vom opri asupra corelației simple. Aceasta poate fi *directă* (atunci cînd pe măsura creșterii variabilei x se mărește și media $\bar{y}$) sau *indirectă* (atunci cînd creșterii caracteristicii x îi corespund descreșteri ale valorilor $\bar{y}$). Indicații asupra gradului de dependență dintre cele două caracteristici luate în considerare se pot obține prin calcularea *coeficientului de corelație* sau a *raportului de corelație*. Primul se calculează pentru regresii liniare, al doilea este folosit în cazul regresiilor neliniare.

6.2. COEFICIENTUL DE CORELAȚIE

Coeficientul de corelație exprimă gradul împrăștierii caracteristicii y în jurul dreptei de regresie, redînd în acest mod intensitatea legăturii dintre x și y. Cu cît această împrăștiere este mai mică cu atît se pot stabili cu mai multă siguranță, pe baza dreptei de regresie, valorile caracteristicii y pentru anumite valori ale caracteristicii x. Mărimea lui este dată de expresia:

$$r = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n\, s_x s_y} \qquad (6.1)$$

unde

r este coeficientul de corelație;
x_i — valorile caracteristicii x;
y_i — valorile caracteristicii y;
n — numărul total al perechilor de unități observate;
s_x — abaterea medie pătratică pentru variabila x;
s_y — abaterea medie pătratică pentru variabila y;
$\bar{x}$ — media valorilor variabilei x;
$\bar{y}$ — media valorilor variabilei y.

Acest raport poate lua valori între -1 și $+1$. Dacă r are valori apropiate de 1 sau -1 se conchide că valorile variabilelor se grupează strîns în jurul dreptelor de regresie. Un $r = +1$ înseamnă o legătură funcțională. Dacă caracteristicile x și y sînt independente, coeficientul de corelație este egal cu 0. Dar, sînt situații cînd coeficientul de corelație este apropiat sau chiar egal cu 0, iar cele două caracteristici sînt corelate (acesta este cazul unor

Scanned with OKEN Scanner

corelații neliniare, cînd trebuie calculat raportul de corelație). Toate legăturile stochastice se caracterizează printr-un coeficient de corelație mai mic decît ± 1.

Înainte de a trece la calculul coeficientului de corelație facem următoarele mențiuni:

a) Indicațiile coeficientului de corelație sînt valabile numai în cazul cînd distribuția bidimensională respectivă este normală sau apropiată de aceasta.

b) La interpretarea coeficientului de corelație se impune o deosebită prudență. Prezența unui coeficient de corelație ridicat între caracteristicile A și B nu trebuie interpretată ca o dovedire a dependenței cauzale a caracteristicii B de caracteristica A. Corelația dintre A și B de multe ori este determinată de alți factori care influențează concomitent asupra ambelor caracteristici. De pildă, corelația ridicată între diametrele și înălțimile arborilor în arboretele echiene nu trebuie interpretată ca o legătură cauzală; ambele caracteristici sînt legate de dezvoltarea generală a biogeocenozei respective. De aceea analiza relațiilor cauzale între caracteristicile luate în studiu trebuie făcută prin prizma metodelor specifice disciplinei respective.

Simpla stabilire a unui coeficient de corelație nu este decît un singur pas, de obicei primul, în cercetarea științifică respectivă. Sarcina ulterioară a investigațiilor științifice este de a descoperi legăturile necesare dintre fenomenele luate în considerare.

c) La calculul coeficientului de corelație trebuie evitată includerea unor unități ce fac parte din alte colectivități statistice. Spre exemplu, este lipsit de sens coeficientul de corelație privind legătura dintre diametre și înălțimi la arboretele bietajate care în prealabil n-au fost separate în elemente de arboret.

d) Eșantionul luat în considerare trebuie ales la întîmplare.

e) Se impune multă prudență la generalizarea coeficienților de corelație stabiliți în anumite condiții particulare. Spre exemplu, coeficientul de corelație dintre creșterea în diametru și grosimea arborilor depinde în mare măsură de vîrstă, consistența și clasa de producție a arboretelor cercetate. Coeficienți de corelație medii se pot stabili numai pentru selecții extrase din aceeași colectivitate generală.

Față de cele menționate este necesar să se aprecieze just conținutul coeficientului de corelație în condițiile în care el poate fi utilizat, cunoscînd că și acest indicator nu are decît o putere de exprimare limitată a relațiilor dintre fenomene.

6.2. CALCULUL COEFICIENTULUI DE CORELAȚIE

Calculul coeficientului de corelație în cazul unui număr redus de observații. Într-un arboret de salcîm, la 20 arbori aleși la întîmplare s-au măsurat atît diametrele la înălțimea pieptului cît și lățimea ultimelor 10 inele anuale (tabelul 6.2).

Scanned with OKEN Scanner

Se pune întrebarea, care este intensitatea corelației dintre diametrul și lățimea inelelor anuale pe baza selecției efectuate. În acest scop vom folosi coeficientul de corelație, deoarece, în baza altor studii, cunoaștem că legătura dintre diametru și lățimea inelului anual este liniară.

Tabelul 6.2

Diametrele și lățimea ultimelor *10* inele anuale la *20* arbori într-un arboret de salcîm

Diametrul cm	Lățimea ultimelor inele anuale mm	Diametrul cm	Lățimea ultimelor inele anuale mm
20	10	14	4
16	7	26	18
18	9	20	9
14	7	20	13
20	12	18	8
24	14	12	6
22	11	20	7
14	6	16	7
16	7	24	16
17	8	22	14

Aplicarea formulei 6.1 este greoaie. Cu mult mai comod se lucrează cu formula analogă:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum_y^2)]}} \qquad (6\ 2a)$$

În cazul valorilor grupate în tabele de corelație, în formula 6.2*a* intervine frecvența perechilor de valori x și y, așa încît vom avea:

$$r = \frac{n\sum_x \sum_y n_{xy}\, xy - \left(\sum_x x n_x\right)\left(\sum_y y n_y\right)}{\sqrt{\left[n\sum_x x^2 n_x - \left(\sum_x n_x x\right)^2\right]\left[n\sum_y y^2 n_y - \left(\sum_y n_y y\right)^2\right]}}$$

În tabelul 6.3 se poate urmări modul de determinare a sumelor necesare. Introducînd sumele obținute în formula (6.2a) rezultă:

$$r = \frac{20 \times 3\,846 - 373 \times 193}{\sqrt{[(20 \times 7\,233 - 373^2)]\ [(20 \times 2\,129 - 193^2)]}} = +\,0{,}91$$

După cum se vede valoarea coeficientului de corelație este foarte mare, fapt explicabil dacă avem în vedere caracteristicile arboretului.

Scanned with OKEN Scanner

Calculul coeficientului de corelaţie în cazul unui număr mare de observaţii. În cazul unei selecţii de volum mare calculele după formula (6.2) sînt destul de complicate şi răpitoare de timp. De aceea în asemenea situaţii se folosesc procedee adecvate care pornesc de la gruparea valorilor observate în *tabele de corelaţie*.

Tabelul 6.3

Calculul sumelor care intervin în formula (6.2 a)

x	y	$x.y$	x^2	y^2
20	10	200	400	100
16	7	112	256	49
18	9	162	324	81
14	7	98	196	49
20	12	240	400	144
24	14	336	576	196
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
373	193	3 846	7 233	2 129
$\sum x$	$\sum y$	$\sum xy$	$\sum x^2$	$\sum y^2$

Pentru asemenea cazuri prezentăm în continuare *metoda momentelor*. Acestea din urmă se calculează după *procedeul sumelor* (§ 2.5).

Tabelul 6.4

Tabel de corelaţie
(Lungimea şi greutatea a *1000* seminţe de salcîm)

Greutatea y_i	lungimea, în mm, x_i							Total
	39	42	45	48	51	54	57	
29					6	6	3	15
26				2	9	3	2	16
23	1		6	25	10	2		44
20	1	8	37	24	11	1		82
17	5	9	10	3	1			28
14	3	3	2					8
Total	10	20	55	54	37	12	5	193

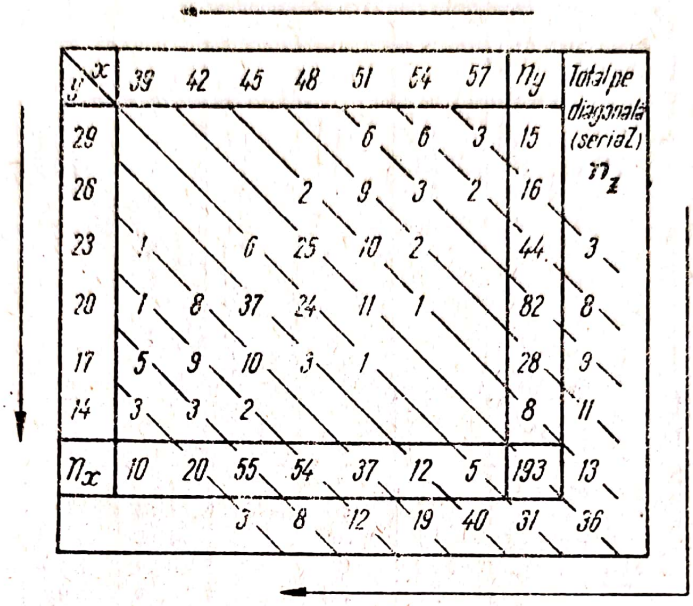
Calculul coeficientului de corelaţie prin procedeul sumelor se rezumă la următoarele operaţii:

a) Datele experimentale se trec în tabelul de corelaţie (tab. 6.4) în care pe primul rînd orizontal se înscriu valorile variabilei independente, în *ordine crescătoare*, iar pe prima coloană — valorile caracteristicii dependente, în ordine *descrescătoare*.

Scanned with OKEN Scanner

b) Se adună frecvențele pe verticale, pe orizontale și pe diagonale. Frecvențele adunate pe verticale formează „seria x", cele adunate pe orizontale formează „seria y", iar cele adunate pe diagonală formează „seria z" (tabelul 6.5).

Tabelul 6.5



c) Frecvențele seriilor x, y și z se trec în tabelul 6.6, în ordinea indicată de săgețile din tabelul 6.5. Se formează o serie pentru control — seria u — adunînd frecvențele pe orizontală. Suma frecvențelor seriei u va trebui să fie egală cu triplul numărului de măsurători ($193 \times 3 = 579$). În continuare datele se prelucrează simultan, potrivit regulilor descrise la § 2.5. Liniile centrale se vor trasa la același nivel pentru toate seriile.

Controlul calculelor constă în aceea că se verifică dacă sumele valorilor ajutătoare a_1, a_2, b_1 și b_2, pentru seriile x, y și z, sînt egale cu valorile corespunzătoare ale seriei u.

O dată stabilită exactitatea calculelor efectuate, în continuare, se determină separat pentru fiecare serie, momentele obișnuite de ordinul I și II, momentul central de ordinul II, media, cît și abaterea medie pătratică. Pentru ultimele două serii, media și abaterea medie pătratică nu se calculează deoarece nu au nici un sens.

d) Se calculează coeficientul de corelație după formula:

$$r = \frac{\mu_z - \mu_x - \mu_y}{2\sqrt{\mu_x \cdot \mu_y}} \qquad (6.3)$$

Potrivit calculelor efectuate în tabelul 6.6, rezultă următoarele valori:

$$\mu_x = 1{,}77; \qquad \mu_y = 1{,}41; \qquad \mu_z = 5{,}50.$$

Tabelul 6.6

Calculul coeficientului de corelaţie prin metoda momentelor

Seria x			Seria y			Seria z			„Seria u" control		
n_i	I	II	n_i	I	II	n_i	I	II	n_i	I	II
						3	3	3	3	3	3
						8	11	14	8	11	14
5	5	5				9	20	34	14	25	39
12	17	22	15	15	15	11	31	65	38	63	102
37	54	76	16	31	46	13	44	109	66	129	231
54	108	—	44	75	—	36	80	—	134	263	—
55	—	—	82	—	—	31	—	—	168	—	—
20	30	—	28	36	—	40	82	—	88	148	—
10	10	10	8	8	8	19	42	79	37	60	97
						12	23	37	12	23	37
						8	11	14	8	11	14
						3	3	3	3	3	3
n 193			193			193			579		
a	184	103		121	61		189	225		494	389
b	40	10		44	8		161	133		245	151
S	224	113		165	69		350	358		739	540
d	144	93		77	53		28	92		249	238

$m_1 = 0{,}75$
$m_2 = 2{,}33$
$\mu_x = 1{,}77$
$s_x = 3\sqrt{1{,}77} = 3{,}99$
$\bar{x} = 47{,}2$

$m_1 = 0{,}40$
$m_2 = 1{,}57$
$\mu_y = 1{,}41$
$s_y = 3\sqrt{1{,}41} = 3{,}56$
$\bar{y} = 21{,}2$

$m_1 = 0{,}145$
$m_2 = 5{,}52$
$\mu_z = 5{,}50$

$r = \dfrac{\mu_z - \mu_x - \mu_y}{2\sqrt{\mu_x \mu_y}} = \dfrac{5{,}50 - 1{,}77 - 1{,}41}{2\sqrt{1{,}77 \cdot 1{,}41}}$

$r = 0{,}73$

Introducînd aceste valori în formula 6.3 obţinem:

$$r = \frac{5{,}50 - 1{,}77 - 1{,}41}{2\sqrt{1{,}77 \times 1{,}41}} = 0{,}73$$

Aşadar, între lungimea şi greutatea seminţelor de salcîm există o corelaţie liniară pozitivă, caracterizată printr-un coeficient de corelaţie de 0,73.

În cazul cînd se dispune de o maşină de calculat, se recomandă următorul procedeu, bazat tot pe însumarea pe diagonală. Coeficientul de corelaţie se calculează după aceeaşi formulă (6.3) unde:

$$\mu_x = \frac{2\Sigma x''}{N} - \frac{\Sigma x'}{N} - \left(\frac{\Sigma x'}{N}\right)^2 \qquad (6.4)$$

$$\mu_y = \frac{2\Sigma y''}{N} - \frac{\Sigma y'}{N} - \left(\frac{\Sigma y'}{N}\right)^2 \qquad (6.5)$$

$$\mu_z = \frac{2\Sigma z''}{N} - \frac{\Sigma z'}{N} - \left(\frac{\Sigma z'}{N}\right)^2 \qquad (6.6)$$

Scanned with OKEN Scanner

BIBLIOTECA Centrală IAȘI

Sumele $\Sigma x'$, $\Sigma x''$, $\Sigma y'$, $\Sigma y''$, $\Sigma z'$, și $\Sigma z''$ rezultă din tabelul 6.8 prin simple însumări, identice celor efectuate în tabelul 6.6, cu excepția că nu se mai delimitează cele două sectoare prin liniuțe de demarcație. Spre exemplu, coloana x' se obține astfel: 2; $2 + 2 = 4$; $4 + 0 = 4$ ș.a.

Tabelul 6.7

Corelația dintre diametre și coeficientul de formă la un arboret echien la molid

Coeficientul de formă f	Diametrul arborilor la înălțimea pieptului (d)											Total y	Total pe diagonale z
	16	20	24	28	32	36	40	44	48	52	56		
0,57	2											2	—
0,54	2	8	6	2	4							22	—
0,51	8	18	12	16	8	8						70	—
0,48		20	28	18	32	18	8	4				128	—
0,45		8	18	22	24	32	18	4				126	—
0,42		2	6	10	8	20	10	6				62	—
0,39		2	2	2	2	4	4				2	18	—
0,36												—	—
0,33					2					2		4	
Total x	12	58	72	70	80	82	40	14	0	2	2	432	6
													24
													52
	—	—	—	—	—	—	—	2	6	16	58	82	88	98

Tabelul 6.8

Calculul sumelor $\Sigma x'$, $\Sigma x''$, $\Sigma y'$, $\Sigma y''$, $\Sigma z'$ și $\Sigma z''$ pentru datele din tabelul 6.7

Seria x			Seria y			Seria z		
x	x'	x''	y	y'	y''	z	z'	z''
2	2	2	2	2	2	6	6	6
2	4	6	22	24	26	24	30	36
0	4	10	70	94	120	52	82	118
14	18	28	128	222	342	98	180	298
40	58	86	126	348	690	88	268	566
82	140	226	62	410	1 100	82	350	916
80	220	446	18	428	1 528	58	408	1 324
70	290	736	—	428	1 956	16	424	1 748
72	362	1 098	4	432	2 388	6	430	2 178
58	420	1 518				2	432	2 610
12	432	1 950						
432	1 950	6 106	432	2 388	8 152	432	2 610	9 800
N	$\Sigma x'$	$\Sigma x''$	N	$\Sigma y'$	$\Sigma y''$	N	$\Sigma z'$	$\Sigma z''$

Scanned with OKEN Scanner

Aplicînd formulele (6.4), (6.5) și (6.6) rezultă:

$$\mu_x = \frac{2\times 6106}{432} - \frac{1950}{432} - \left(\frac{1950}{432}\right)^2 = 2{,}378$$

$$\mu_y = \frac{2\times 8150}{432} - \frac{2388}{432} - \left(\frac{2388}{432}\right)^2 = 1{,}644$$

$$\mu_z = \frac{2\times 9800}{432} - \frac{2610}{432} - \left(\frac{2610}{432}\right)^2 = 2{,}822$$

Introducînd aceste valori în formula (6.3) obținem:

$$r = \frac{\mu_z - \mu_x - \mu_y}{2\sqrt{\mu_x \mu_y}} = \frac{2{,}822 - 2{,}378 - 1{,}644}{2\sqrt{2{,}378\times 1{,}644}} = -0{,}30$$

6.3. STABILIREA SEMNIFICAȚIEI UNEI CORELAȚII

Coeficientul de corelație o dată calculat, este necesar să se stabilească dacă valoarea obținută este reală sau se datorește erorilor întîmplătoare de selecție.

În acest scop, deosebim mai multe situații.

În cazul cînd *volumul selecției este ridicat* și r nu prea mare ($r < 0{,}8$), pentru verificarea semnificației coeficientului de corelație se folosește criteriul u:

$$u = \frac{r}{s_r} \tag{6.7}$$

unde s_r se determină cu ajutorul relației:

$$s_r = \frac{1 - r^2}{\sqrt{N}} \tag{6.9}$$

Valorile lui u se găsesc în anexa 3. Dacă u experimental este mai mare decît u tabelar, la nivelul de semnificație ales, rezultă că legătura celor două caracteristici există în realitate; în caz contrar, rezultatele cercetării nu sînt concludente pentru a susține existența corelației.

Astfel, pentru exemplul din tabelul 6.7 a rezultat un coeficient de corelație $r = -0{,}30$. Eroarea acestui coeficient este:

$$s_r = \frac{1 - 0{,}09}{\sqrt{432}} = 0{,}044$$

Deci

$$u = \frac{0{,}30}{0{,}044} = 6{,}8$$

Pentru nivelul de semnificație de 0,05 (5%), găsim în anexa 3, $u = 1{,}96$, iar la nivelul de semnificație de 0,001 (0,1%) stabilim $u = 3{,}29$. Așadar, putem într-adevăr să afirmăm cu o probabilitate de 99,9%, că între diame-

Scanned with OKEN Scanner

trul arborilor și coeficientul de formă f există o legătură de corelație. Legătura este însă relativ slabă ($r = -0,30$). Pentru probabilitatea de 99,9% intervalul de încredere al coeficientului de corelație teoretic este:

$$r \pm us_r = -0,30 \pm 3,29 \times 0,044 = -0,30 \pm 0,14 \begin{cases} -0,16 \\ -0,44 \end{cases}$$

În cazul unor selecții de volum mic cu un r *mare* distribuția de eșantionaj a coeficienților de corelație nu mai este normală, ceea ce nu mai permite folosirea curbei normale în forma ei clasică. Pentru a nu renunța la avantajele distribuției normale, FISHER a introdus transformarea:

$$z = \frac{1}{2} ln \frac{1+r}{1-r} \qquad (6.9)$$

care are proprietatea de a se distribui normal.

Determinarea lui z în funcție de r se poate face după tabelul 6.9, iar valorile lui r în funcție de z se calculează după tabelul 6.10.

Tabelul 6.9

Valorile funcției $z = \frac{1}{2} ln \frac{1+r}{1-r}$ de la $r = 0,00$ la $r = 0,999$

r	0	1	2	3	4	5	6	7	8	9
0,00	0,0000	0,0101	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
1	0,1003	0,1104	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
3	0,3095	0,3205	0,3316	0,3428	0,3541	0,3654	0,3767	0,3884	0,4001	0,4118
4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5239	0,5361
5	0,5493	0,5627	0,5764	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
6	0,6932	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
7	0,8673	0,8872	0,9077	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
9	1,4722	1,5275	1,5890	1,6584	1,7381	1,8318	1,9459	2,0923	2,2976	2,6467
0,999	2,6466	2,6996	2,7587	2,8257	2,9031	2,9945	3,1063	3,2504	3,4534	3,8002

Eroarea standard a parametrului z este dată de formula:

$$s_z = \frac{1}{\sqrt{n-3}}.$$

E x e m p l u. Pentru datele din tabelul 6.2, ($r = +0,91$; $n = 20$) obținem:

$$z = \frac{1}{2} ln \frac{1+0,91}{1-0,91} = 1,5275; \qquad s_z = \frac{1}{\sqrt{20-3}} = 0,242.$$

Pentru nivelul de semnificație de 5% ($u = 1,96$ — anexa 3) obținem următorul interval de încredere pentru parametrul z:

$$1,527 \pm 1,96 \times 0,242 = 1,527 \pm 0,475 \begin{cases} 1,052 \\ 2,002 \end{cases}$$

Scanned with OKEN Scanner

Tabelul 6.10

Valorile coeficientului de corelaţie r, corespunzătoare diferitelor valori ale funcţiei z a a lui Fisher de la z = 0,01 la z = 3,00

z	,01	,02	,03	,04	,05	,06	,07	,08	,09	0,10
0,0	0,0100	0,0200	0,0300	0,0400	0,0500	0,0599	0,0699	0,0798	0,0898	0,0997
0,1	0,1096	0,1194	0,1293	0,1391	0,1489	0,1586	0,1684	0,1781	0,1877	0,1974
0,2	0,2070	0,2165	0,2260	0,2355	0,2449	0,2543	0,2636	0,2729	0,2821	0,2913
0,3	0,3004	0,3095	0,3185	0,3275	0,3364	0,3452	0,3540	0,3627	0,3714	0,3800
0,4	0,3885	0,3969	0,4053	0,4136	0,4219	0,4301	0,4382	0,4462	0,4542	0,4621
0,5	0,4699	0,4777	0,4854	0,4930	0,5005	0,5080	0,5154	0,5227	0,5299	0,5370
0,6	0,5441	0,5511	0,5580	0,5649	0,5717	0,5784	0,5850	0,5915	0,5980	0,6044
0,7	0,6107	0,6169	0,6231	0,6291	0,6351	0,6411	0,6469	0,6528	0,6584	0,6640
0,8	0,6696	0,6741	0,6805	0,6858	0,6911	0,6963	0,7014	0,7064	0,7114	0,7163
0,9	0,7211	0,7259	0,7306	0,8352	0,7398	0,7443	0,7487	0,7531	0,7574	0,7616
1,0	0,7658	0,7699	0,7739	0,7779	0,7818	0,7857	0,7895	0,7932	0,7969	0,8005
1,1	0,8041	0,8076	0,8110	0,8144	0,8178	0,8210	0,8243	0,8275	0,8306	0,8337
1,2	0,8367	0,8397	0,8426	0,8455	0,8483	0,8511	0,8538	0,8565	0,8591	0,8617
1,3	0,8643	0,8668	0,8692	0,8717	0,8741	0,8764	0,8787	0,8810	0,8832	0,8854
1,4	0,8675	0,8896	0,8917	0,8937	0,8957	0,8977	0,8966	0,9015	0,9033	0,9051
1,5	0,9069	0,9087	0,9104	0,9121	0,9138	0,9154	0,9170	0,9186	0,9201	0,9217
1,6	0,9232	0,9246	0,9261	0,9275	0,9289	0,9302	0,9316	0,9329	0,9341	0,9354
1,7	0,9366	0,9379	0,9391	0,9402	0,9414	0,9425	0,9436	0,9447	0,9458	0,94681
1,8	0,94783	0,94884	0,94983	0,95080	0,95175	0,95268	0,95359	0,95449	0,95537	0,95624
1,9	0,95609	0,95792	0,95873	0,95953	0,96032	0,96109	0,96185	0,96259	0,96331	0,96403
2,0	0,96473	0,96541	0,96609	0,96675	0,96739	0,96803	0,96865	0,96926	0,96986	0,97045
2,1	0,97103	0,97159	0,97215	0,97269	0,97323	0,97375	0,97426	0,97477	0,97526	0,97574
2,2	0,97622	0,97668	0,97714	0,97759	0,97803	0,97846	0,97888	0,97929	0,97970	0,98010
2,3	0,98049	0,98087	0,98124	0,98161	0,98197	0,98233	0,98267	0,98301	0,98335	0,98367
2,4	0,98399	0,98431	0,98462	0,98492	0,98522	0,98551	0,98579	0,98607	0,98635	0,98661
2,5	0,98688	0,98714	0,98739	0,98764	0,98788	0,98812	0,98835	0,98858	0,98881	0,98903
2,6	0,98924	0,98945	0,98966	0,98987	0,99007	0,99026	0,99045	0,99064	0,99083	0,99101
2,7	0,99118	0,99136	0,99153	0,99170	0,99186	0,99202	0,99218	0,99233	0,99248	0,99263
2,8	0,99278	0,99292	0,99306	0,99320	0,99333	0,99346	0,99359	0,99372	0,99384	0,99396
2,9	0,99408	0,99420	0,99431	0,99443	0,99454	0,99464	0,99475	0,99485	0,99495	0,99505

Trecînd acum de la z la r, după tabelul 6.10, obţinem:

— pentru $z_1 = 1{,}052$; $r = 0{,}78$

— pentru $z_2 = 2{,}002$; $r = 0{,}96$

Coeficientul de corelaţie al colectivităţii generale, din care a fost extras eşantionul respectiv, este cuprins deci în limitele 0,78 — 0,96. Acest lucru se afirmă cu o probabilitate de 0,95.

Uneori, determinînd valoarea coeficientului de corelaţie dorim să cunoaştem dacă pentru colectivitatea generală *r este mai mare decît 0*. Se pune deci întrebarea dacă, în colectivitatea respectivă există sau nu o corelaţie semnificativă între cele două caracteristici luate în considerare.

În acest scop se determină valoarea minimă pe care trebuie să o aibă un coeficient de corelaţie al unui eşantion pentru a putea fi considerat semnificativ la un anumit nivel de semnificaţie.

Scanned with OKEN Scanner

Tabelul 6.11

Limita inferioară a coeficientului de corelație semnificativ

Numărul gradelor de libertate	Nivelul de semnificație				
	0,10	0,05	0,02	0,01	0,001
3	0,805	0,878	0,9343	0,9587	0,9911
4	0,729	0,811	0,882	0,9172	0,9741
6	0,621	0,707	0,789	0,834	0,9249
8	0,549	0,632	0,715	0,765	0,872
10	0,497	0,576	0,658	0,708	0,823
12	0,457	0,532	0,612	0,661	0,780
14	0,426	0,497	0,574	0,623	0,742
16	0,400	0,468	0,543	0,590	0,708
18	0,378	0,444	0,516	0,561	0,679
20	0,360	0,423	0,492	0,537	0,652
25	0,323	0,381	0,445	0,487	0,597
30	0,296	0,349	0,409	0,449	0,554
35	0,275	0,325	0,381	0,418	0,519
40	0,257	0,304	0,358	0,393	0,490
45	0,243	0,288	0,338	0,372	0,465
50	0,231	0,273	0,322	0,354	0,443
60	0,211	0,250	0,295	0,325	0,408
70	0,195	0,232	0,274	0,302	0,380
80	0,183	0,217	0,257	0,283	0,357
90	0,173	0,205	0,242	0,267	0,338
100	0,164	0,195	0,230	0,254	0,321

Pentru eșantioane cu un efectiv mare ($n > 30$) această valoare „critică" a coeficientului de corelație este dată de relația:

$$r = \frac{u}{\sqrt{n-1}} \tag{6.10}$$

unde u se determină după tabelul din anexa 3, în raport cu nivelul de semnificație ales. Astfel, pentru un nivel de semnificație de 0,05 ($u = 1,96$), pentru $n = 100$, avem

$$r = \frac{1,96}{\sqrt{100-1}} = 0,197$$

Orice coeficient mai mare decît 0,197 este semnificativ (orice r mai mic decît 0,197 nu este semnificativ).

În cazul unor selecții de volum mic, aplicăm criteriul t după formula:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{6.11}$$

Potrivit acestei formule s-a întocmit tabelul 6.11 care arată valorile coeficientului de corelație r (la diferite nivele de semnificație) ce trebuie depășite pentru a accepta ipoteza existenței corelației dintre cele două caracteristici. Numărul gradelor de libertate este dat de relația $f = n - 2$.

Scanned with OKEN Scanner

Dacă r experimental este mai mic decît r tabelar, nu se poate susține existența corelației. Pentru datele din tabelul 6.2 ($r = +0,91$; $n = 20$; $f = 18$) obținem o valoare tabelară de 0,444, la nivelul de semnificație 0,05 (5%). Valoarea obținută este mult inferioară coeficientului de corelație experimental. Aceeași concluzie se menține și la nivelul de semnificație de 0,001.

Prin urmare, corelația dintre diametre și lățimea ultimelor 10 inele anuale la arboretul de salcîm studiat este dovedită.

Pentru verificarea egalității a doi coeficienți de corelație se folosește funcția lui Fisher:

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}, \tag{6.9}$$

Se stabilește diferența dintre z_1 și z_2, corespunzătoare celor doi coeficienți de corelație; diferența obținută se compară cu eroarea diferenței dintre z_1 și z_2. Această eroare este dată de expresia:

$$s_{z_1} - s_{z_2} = \sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}} \tag{6.12}$$

Dacă

$$|z_1 - z_2| < 2s_{z_1 - z_2} \tag{6.13}$$

înseamnă că diferența dintre cei doi coeficienți de corelație este nesemnificativă (pentru nivelul de semnificație de 0,05)*. În acest caz coeficienții de corelație empirici r_1 și r_2 se consideră estimații ale unuia și aceluiași coeficient de corelație. Dacă

$$|z_1 - z_2| > 2s_{z_1 - z_2} \tag{6.14}$$

rezultă că cei doi coeficienți de corelație se referă la colectivități statistice distincte.

Exemplul 1. La două arborete de molid din clasa a IV-a de vîrstă s-au măsurat creșterile în diametru și grosimea la înălțimea pieptului. La primul arboret s-au efectuat măsurători la 150 arbori, rezultînd un coeficient de corelație de 0,720, iar la al doilea arboret au fost măsurați 180 arbori pentru care s-a stabilit un coeficient de corelație de 0,740 (Giurgiu, 1957). Se pune întrebarea dacă acești doi coeficienți de corelație se deosebesc semnificativ. Aplicînd formula (6.9) sau tabelul 6.9 obținem:

$$z_1 = 0,9077; \qquad z_2 = 0,9505$$

$$z_2 - z_1 = 0,9505 - 0,9077 = 0,0428$$

După formula (6.12) avem:

$$s_{z_2 - z_1} = \sqrt{\frac{1}{150-3} + \frac{1}{180-3}} = 0,112$$

* Pentru alte niveluri de semnificație se va folosi anexa 3. Astfel, pentru nivelul de semnificație de 0,01 condiția este:

$$|z_1 - z_2| < 2,58\, s_{z_1 - z_2}$$

Întrucît $z_2 - z_1 = 0{,}0428$ este mai mică decît $2s_{z_2 - z_1} = 0{,}224$, rezultă că diferența dintre cei doi coeficienți de corelație este neînsemnată. Conchidem deci că acești coeficienți de corelație sînt estimații ale unuia și aceluiași coeficient de corelație general, diferența semnalată fiind întîmplătoare (pentru nivelul de semnificație de 0,05).

În cercetările biometrice, pe baza unor selecții diferite se calculează mai mulți coeficienți de corelație. Desigur, între coeficienții de variație calculați sînt abateri mai mici sau mai mari. De aceea se pune *problema verificării ipotezei egalității acestor coeficienți de corelație.*

Problema pusă se rezolvă cu ajutorul criteriului χ^2 (anexa 6) folosind $f = K - 1$ grade de libertate (K fiind numărul coeficienților de corelație luați în considerare).

Tabelul 6.12

Coeficientul de corelație	Numărul perechilor de valori, observate
0,36	201
0,62	170
0,42	175
0,35	195

E x e m p l u l 2. La 4 arborete s-a stabilit corelația dintre mărimea categoriilor de diametre și procentul creșterii în suprafață de bază, rezultînd valorile din tabelul 6.12.

Urmează să verificăm dacă cei patru coeficienți de corelație sînt estimații ale aceluiași coeficient de corelație teoretic. În acest scop calculăm criteriul χ^2 după formula:

$$\chi^2 = \Sigma (n_i - 3) z_i^2 - \frac{[\Sigma (n_i - 3) z_i]^2}{\Sigma (n_i - 3)} \tag{6.15}$$

unde

$$z_i = \frac{1}{2} ln \frac{1 + r_i}{1 - r_i} \tag{6.16}$$

Desfășurarea calculelor se poate urmări în tabelul (6.13).

Tabelul 6.13

Verificarea ipotezei egalității mai multor coeficienți de corelație

Coeficientul de corelație empiric r_i	Numărul perechilor de valori observate n_i	z_i	$n_i - 3$	$(n_i - 3) z_i$	$(n_i - 3) z_i^2$
0,36	201	0,3767	198	74,6	28,1
0,62	170	0,7250	167	121,1	87,8
0,42	175	0,4477	172	77,0	34,5
0,35	195	0,3654	192	70,2	25,6
Σ	—	—	729	342,9	176,0

$$\chi^2 = 176{,}0 - \frac{(342{,}9)^2}{729} = 14{,}7, \qquad f = K - 1 = 4 - 1 = 3$$

$\chi^2_{tab} = 7{,}81$ (pentru nivelul de semnificație de 0,05);

χ^2 *tab* $<$ χ^2 experimental.

Din tab. 6.13 rezultă $\chi^2 = 14{,}7$.

Scanned with OKEN Scanner

Din anexa 6, la $f = 4 - 1 = 3$ grade de libertate, pentru nivelul de semnificaţie de 0,05, găsim $\chi^2 = 7{,}81$.

Deoarece valoarea calculată $\chi^2 = 14{,}7$ este mai mare decît valoarea tabelară $\chi^2 = 7{,}81$, nu există nici un temei să se accepte ipoteza că cei patru coeficienţi de corelaţie empirici sînt estimaţii ale aceluiaşi coeficient de corelaţie teoretic.

Seria acestor coeficienţi de corelaţie se omogenizează dacă excludem valoarea $r = 0{,}62$. Într-adevăr, în această nouă situaţie χ^2 calculat este 0,7, ia χ^2 tabelar, luat la 2 grade de libertate şi la nivelul de semnificaţie de 0,05, este de 5,99, ceea ce arată că cei trei coeficienţi de corelaţie rămaşi (0,36; 0,42 şi 0,35) pot fi consideraţi ca estimaţii ale aceluiaşi coeficient de corelaţie teoretic. Pentru determinarea coeficientului de corelaţie mediu calculăm iniţial $\bar{z}$ după formula:

$$\bar{z} = \frac{\Sigma (n_i - 3) z_i}{\Sigma (n_i - 3)} = 0{,}393$$

La $\bar{z} = 0{,}393$, după tabelul 6.10, stabilim $\bar{r} = 0{,}37$.

6.4. CORELAŢIA MULTIPLĂ

În cazul influenţei comune a mai multor factori asupra unei caracteristici date, se aplică teoria *corelaţiei multiple*. Intensitatea corelaţiei multiple se exprimă cu ajutorul *coeficientului de corelaţie multiplă* R. Mai simplu este cazul *corelaţiei multiple liniare*.

Pentru corelaţia liniară cu trei caracteristici (y, x şi z) coeficientul de corelaţie multiplă este dat de relaţia:

$$R = \sqrt{\frac{r_{xy}^2 + r_{yz}^2 - 2 r_{xy} \cdot r_{yz} \cdot r_{xz}}{1 - r_{xz}^2}} \qquad (6.17)$$

unde: r_{xy}, r_{yz} şi r_{xz} reprezintă *coeficienţii de corelaţie simplă*.

Coeficientul de corelaţie multiplă are totdeauna semn pozitiv şi este mai mare decît coeficienţii de corelaţie simplă luaţi în parte. Se demonstrează că valoarea lui R se apropie de 1 atunci cînd se include în analiză un număr din ce în ce mai mare de caracteristici factoriale.

Pentru exemplificare, vom stabili intensitata corelaţiei dintre volumul (y), diametrul (x) şi înălţimea (z) unor arbori extraşi dintr-un arboret de molid. Calculele ajutătoare se pot urmări în tabelul 6.14.

În primul rînd se calculează coeficienţii de corelaţie simplă. În acest scop folosim formula (6.1).

Avînd în vedere relaţia (6.17) obţinem coeficientul de corelaţie multiplă $R_{yxz} = 0{,}993$.

Coeficientul obţinut, $R = 0{,}993$, arată intensitatea legăturii reciproce a celor două caracteristici factoriale (diametru x şi înălţimea y) asupra caracteristicii rezultative, volumul arborilor y. Mărimea coeficientului $R = 0{,}993$ dovedeşte o dependenţă de corelaţie multiplă foarte puternică, fapt explicabil dacă avem în vedere natura fenomenului studiat.

Scanned with OKEN Scanner

Tabelul 6.14

Determinarea coeficientului de corelație dintre volumul, diametrul și lungimea fusului la un lot de 10 arbori de molid

x	z	y	$x-\bar{x}$	$z-\bar{z}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(z-\bar{z})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(z-\bar{z})$	$(x-\bar{x})(y-\bar{y})$	$(z-\bar{z})(y-\bar{y})$
15,5	14,1	0,14	—3,05	—2,1	—0,149	9,30	4,41	0,022	6,41	0,508	0,313
20,5	18,5	0,52	+1,95	+2,3	+0,231	3,76	5,29	0,053	4,49	0,450	0,531
12,0	10,8	0,07	—6,55	—5,4	—0,219	42,90	29,16	0,046	37,37	1,326	1,183
12,5	12,4	0,08	—6,05	—3,8	—0,209	36,60	14,44	0,044	22,99	1,264	0,794
16,5	16,1	0,19	—2,05	—0,1	—0,099	4,20	0,01	0,010	0,21	0,205	0,010
24,5	19,2	0,45	+5,95	+3,0	+0,161	35,48	9,00	0,026	17,85	0,953	0,486
24,5	26,6	0,48	+5,95	+4,4	+0,191	35,48	19,36	0,036	26,18	1,136	0,840
23,5	20,6	0,49	+4,55	+4,4	+0,201	23,50	19,36	0,040	21,78	0,995	0,884
12,0	10,9	0,06	—6,55	—5,3	—0,229	42,90	26,09	0,032	34,72	1,500	1,234
24,0	18,8	0,41	+5,45	+2,6	+0,121	29,70	6,76	0,015	14,17	0,659	0,315
185,5	162,0	2,89	0	0	0	263,82	133,88	0,324	186,17	8,996	6,567

$\bar{x} = 18{,}55 \qquad s_x = 5{,}14 \qquad r_{xy} = \dfrac{8{,}996}{10 \times 5{,}14 \times 0{,}18} = 0{,}97$

$\bar{z} = 16{,}2 \qquad s_z = 3{,}66 \qquad r_{yz} = \dfrac{6{,}567}{10 \times 0{,}18 \times 3{,}66} = 0{,}99$

$\bar{y} = 0{,}289 \qquad s_y = 0{,}18 \qquad r_{xz} = \dfrac{186{,}17}{10 \times 5{,}14 \times 3{,}66} = 0{,}99$

$$R_{yxz} = \sqrt{\frac{0{,}97^2 + 0{,}99^2 - 2 \times 0{,}97 \times 0{,}99 \times 0{,}99}{1 - 0{,}99^2}} = 0{,}993$$

Scanned with OKEN Scanner

6.5. CORELAȚIA PARȚIALĂ

În cazul corelației multiple, de multe ori, se impune examinarea corelației dintre două caracteristici în ipoteza că ceilalți factori rămîn constanți și nu influențează asupra legăturii considerate. În acest scop se calculează *coeficienții de corelație parțială*. Aceștia se determină pe baza coeficienților de corelație simpli. În situația a trei caracteristici, se utilizează formula:

$$r_{xy \cdot z} = \frac{r_{xy} - r_{yz} \cdot r_{xz}}{\sqrt{(1 - r_{yz}^2)\ (1 - r_{xz}^2)}} \tag{6.19}$$

Coeficientul $r_{xy \cdot z}$ reprezintă coeficientul de corelație parțială dintre caracteristicile y și x în ipoteza eliminării influenței caracteristicii z asupra legăturii luate în studiu.

În mod analog se pot determina coeficienții de corelație parțială $r_{yz \cdot x}$ și $r_{xz \cdot y}$. Pentru $r_{yz \cdot x}$ avem expresia:

$$r_{yz \cdot x} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1 - r_{xy}^2)\ (1 - r_{xz}^2)}} \tag{6.20}$$

Pentru $r_{xz \cdot y}$ avem:

$$r_{xz \cdot y} = \frac{r_{xz} - r_{yz} \cdot r_{xy}}{\sqrt{(1 - r_{yz}^2)\ (1 - r_{xy}^2)}} \tag{6.21}$$

Pentru exemplul luat la § 6.4, privind corelația multiplă dintre x, y și z, obținem:

$$r_{xy \cdot z} = \frac{0{,}97 - 0{,}99 \times 0{,}99}{\sqrt{(1 - 0{,}99^2)\ (1 - 0{,}99^2)}} = -0{,}50$$

$$r_{yz \cdot x} = \frac{0{,}99 - 0{,}97 \times 0{,}99}{\sqrt{(1 - 0{,}97^2)\ (1 - 0{,}99^2)}} = +0{,}85$$

$$r_{xz \cdot y} = \frac{0{,}99 - 0{,}99 \times 0{,}97}{\sqrt{(1 - 0{,}99^2)\ (1 - 0{,}97^2)}} = +0{,}85$$

6.6. CORELAȚIA RANGURILOR

În paragrafele anterioare s-a tratat problema corelației pentru distribuțiile statistice normale. Sînt însă cazuri cînd tipul distribuției este necunoscut sau este dificil de stabilit din cauza numărului relativ redus de valori observate. În asemenea cazuri metodele clasice devin inaplicabile. În schimb, se pot folosi *metode neparametrice* care fac abstracție de condiția cunoașterii prealabile a tipului de curbă a distribuției. Intensitatea corelației este dată de *coeficientul de corelație a rangurilor* (S n e d e c o r, 1957). După S p e a r m a n acesta este dat de relația:

$$r_s = 1 - \frac{6 \Sigma d^2}{n (n^2 - 1)} \tag{6.22}$$

Scanned with OKEN Scanner

unde:

d este fluctuația de rang, adică diferența dintre rangul după o variabilă și rangul după cealaltă variabilă.

E x e m p l u. La 8 arborete de stejar de aceeași proveniență, de aceeași vîrstă și dezvoltate în condiții staționale asemănătoare, s-a determinat atît consistența cît și procentul arborilor în curs de uscare. În urma observațiilor efectuate au rezultat datele din tabelul 6.15, unde în coloanele 3 și 4 se formează două șiruri de numere aranjate după rangul lor. În penultima coloană s-a făcut diferența dintre ranguri, iar în ultima coloană diferențele obținute s-au ridicat la pătrat. Aplicînd formula (6.22) rezultă un coeficient de corelație însemnat (— 0,81), ceea ce arată că procentul arborilor în curs de uscare este puternic corelat cu consistența arboretelor.

Tabelul 6.15

Calculul coeficientului de corelație al rangurilor (corelația dintre consistența arboretelor de stejar și procentul arborilor în curs de uscare)

Consistența arboretelor	Procentul arborilor în curs de uscare	Rangul după consistența arboretelor	Rangul după procentul arborilor în curs de uscare	Diferența	
				d	d^2
0,3	22	1	5	—4	16
0,4	26	2	8	—6	36
0,5	24	3	6	—3	9
0,6	25	4	7	—3	9
0,7	20	5	4	+1	1
0,8	17	6	3	+3	9
0,9	10	7	1	+6	36
1,0	15	8	2	+6	36
Σ	—	—	—		152

$$r_s = 1 - \frac{6 \Sigma d^2}{n(n^2-1)} = 1 - \frac{6 \times 152}{8(8^2-1)} = -0{,}81$$

Pentru determinarea coeficientului de corelație a rangurilor se pot folosi și alte procedee tot atît de simple, cum sînt cele elaborate de K e n d a l l, H o e f f d i n g sau T u k e y (M u n t e a n u, 1964).

6.7. TABELE DE CONTINGENȚĂ

În acest paragraf vom căuta să stabilim prezența legăturii dintre caracteristici în cazul atributelor nemăsurabile, cîntărind concordanțele cu discordanțele prin intermediul frecvențelor caracteristicilor luate în studiu. Cazul cel mai simplu este acela cînd se iau în considerare numai două atribute.

Luăm următoarele exemple:

Exemplul 1. Într-un arboret de stejar cu gîrniță, în vîrstă de 140 ani, s-au găsit 15 exemplare uscate dintre care 5 de gîrniță și 10 de stejar. În rest exemplarele sănătoase erau repartizate astfel: gîrniță 60, stejar 25. Se cere să se stabilească dacă fenomenul de uscare este corelat cu specia.

În acest scop frecvențele obținute le înregistrăm într-un tabel de contingență (tabelul 6.16).

Într-o formă generală tabelul prezentat se scrie astfel:

Tabelul 6.16

	Specia		Total
	Gîrniță	Stejar	
Arbori sănătoși	60	25	85
Arbori uscați	5	10	15
Total	65	35	100

Tabelul 6.17

Atributul B	Atributul A		Total
	A_1	A_2	
B_1	a	b	$a+b$
B_2	c	d	$c+d$
Total	$a+c$	$b+d$	$a+b+c+d$

Pentru a scoate în evidență existența asociației se poate folosi testul (criteriul) χ^2, calculat după formula:

$$\chi^2 = \frac{\left\{\left|ad-bc\right|-\frac{1}{2}n\right\}^2}{(a+b)\ (c+d)\ (a+c)\ (b+d)} \tag{6.23}$$

Valoarea experimentală a lui χ^2 se compară cu valoarea tabelară, luată din anexa 6, *folosind un singur grad de libertate.*

Pentru exemplul considerat obținem:

$$\chi^2 = \frac{100\left\{\left|60\times10-25\times5\right|-\frac{100}{2}\right\}^2}{(60+25)\ (5+10)\ (60+5)\ (25+10)} = 6{,}2$$

După tabelul din anexa 6 la $f=1$, pentru nivelul de semnificație de 0,05, rezultă $\chi^2 = 3{,}84$. Deoarece χ^2 experimental este mai mare decît χ^2 tabelar, conchidem că între fenomenul de uscare și specie există o asociație semnificativă. Deci, în condițiile staționale date, gîrnița este mai rezistentă la uscare în comparație cu stejarul.

Un alt caz, frecvent întîlnit, este acela cînd valorile observate se pot grupa într-un *tabel de contingență cu 2 rînduri* și *c coloane*, avînd forma din tabelul 6.18.

Tabelul 6.18

. . .	a_i	A
. . .	b_i	B
. . .	n_i	N

În acest tabel a_i reprezintă frecvențele trecute pe primul rînd, iar b_i frecvențele trecute pe al doilea rînd. Prin n_i sînt însemnate sumele frecvențelor pe verticală. Sumele pe orizontală sînt reprezentate, respectiv, prin A, B și N.

Scanned with OKEN Scanner

Valoarea experimentală a criteriului χ^2 este dată de formula Brandt-Snedecor:

$$\chi^2 = \frac{\sum \frac{a_i^2}{n_i} - \frac{A^2}{N}}{K(1-K)} = \frac{\sum \frac{b_i^2}{n_i} - \frac{B^2}{N}}{K(1-K)} \qquad (6.24)$$

unde:

$$K = \frac{A}{N} \; ; \qquad 1 - K = \frac{B}{N}$$

iar numărul gradelor de libertate

$$f = c - 1 \qquad (6.25)$$

în care c este numărul coloanelor.

Exemplul 2. La 120 arborete de molid de aceeaşi vîrstă şi proveniență, grupate în 6 tipuri de pădure, s-a studiat frecvența doborîturilor produse de vînt, rezultînd următorul tabel de contingență ($c = 6$) (tabelul 6.19)

Tabelul 6.19

Prezența doborîturilor produse de vînt	Tipul de pădure						Total
	I	II	III	IV	V	VI	
Nu	14	8	7	15	9	20	73
Da	10	10	8	6	5	8	47
Total	24	18	15	12	14	28	120

Se pune întrebarea în ce măsură prezența doborîturilor produse de vînt se asociază cu tipul de pădure?

Calculînd valoarea χ^2 obținem:

$$\chi^2 = \frac{\frac{10^2}{24} + \frac{10^2}{18} + \frac{8^2}{15} + \ldots + \frac{8^2}{28} - \frac{47^2}{120}}{\frac{73}{120} \times \frac{47}{120}} = 6{,}0$$

Valoarea tabelară a criteriului χ^2, luată din anexa 6 la $6 - 1 = 5$ grade de libertate, este de 11,07.

Conchidem deci că nu trebuie respinsă ipoteza omogenității distribuțiilor studiate, deoarece χ^2 tabelar pentru nivelul de semnificație de 5% este mai mare decît χ^2 calculat. Prin urmare, în baza datelor prezentate,

Scanned with OKEN Scanner

nu se poate susține existența unei asociații între prezența doborîturilor produse de vînt și tipul de pădure.

În cazul unui tabel de contingență *cu c coloane și r rînduri* (tabelul 6.20) se folosește formula

Tabelul 6.20

Tabelul de asociație r × c

a, b	A_i
c, d	.
.	.
.	.
.	.
B_i	N

$$\chi^2 = \sum \frac{(O - E)^2}{E} \qquad (6.26)$$

unde:

O sînt frecvențele statistice;

E — frecvențele „teoretice"

Frecvențele „teoretice" sînt date de formula

$$E_i = \frac{(A_i)\ (B_i)}{N} \qquad (6.27)$$

Numărul gradelor de libertate rezultă din relația:

$$f = (r - 1)\ (c - 1) \qquad (6.28)$$

Exemplul 3. Într-un arboret de fag, de 80 ani, s-a efectuat o inventariere a arborilor pe clase de calitate, separat pe clase de creștere (după poziția arborilor în arboret) (tabelul 6.21).

Se pune întrebarea în ce măsură clasa de calitate este asociată cu clasa de creștere a arborilor? În acest scop folosim criteriul χ^2, calculat după formula (6.26). Frecvențele „teoretice" E_i sînt calculate în tabelul 6.22. (Prima frecvență teoretică s-a calculat, după formula 6.27, astfel: 25 × × 73 : 150 = 12)

Aplicînd formula (6.26), rezultă:

$$\chi^2 = \frac{(10 - 12)^2}{12} + \frac{(6-6)^2}{6} + \frac{(9-7)^2}{7} \cdots \frac{(13-18)^2}{18} = 5{,}9$$

Numărul gradelor de libertate va fi:

$$f = (3 - 1)\ (3 - 1) = 4$$

Tabelul 6.21

Repartiția arborilor pe clase de calitate și clase de creștere, într-un arboret echien de fag

Clasa de calitate	Clasa de creștere			Total
	I + II	III	IV	
IV	10	6	9	25
III	25	15	20	60
I + II	38	14	13	65
Total	73	35	42	150

Tabelul 6.22

Frecvențele teoretice pentru datele din tabelul 6.21

Clasa de calitate	Clasa de creștere		
	I + II	III	IV
IV	12	6	7
III	29	14	17
I + II	32	15	18

Scanned with OKEN Scanner

Valoarea tabelară a criteriului χ^2 pentru nivelul de semnificație de 0,05 este de 9,49. Întrucît valoarea calculată a criteriului χ^2 este mai mică decît valoarea tabelară a acestuia, conchidem că legătura dintre clasa de calitate și clasa de creștere a arborilor nu este dovedită. (Nu există temei pentru respingerea ipotezei nule.)

Formulele prezentate mai sus sînt aplicabile numai în cazul cînd în tabelul de contingență nu apar frecvențe „teoretice" mai mici de 5. În caz contrar, fie se trece la restrîngerea numărului de coloane sau rînduri, fie se aplică procedee speciale (B a i l e y, 1959; S t e i n b a c h, 1961; V e s s e r e a u, (1960), ș.a.)

* * *

Testul χ^2 este frecvent folosit în cercetările genetice la analiza frecvențelor în cazul disjuncțiunii caracterelor în generația a două hibridă și în celelalte generații hibride precum și în multe alte împrejurări.

Scanned with OKEN Scanner

VII

EXEMPLE DE ECUAȚII DE REGRESIE DIN BIOMETRIA FORESTIERĂ

7.1 GENERALITĂȚI

S-a arătat (§ 6.1) că *ecuațiile de regresie* se deosebesc esențial de ecuațiile rigide (funcționale), prin aceea că primele dau previziuni aproximative referitoare la medii, cîtă vreme ecuațiile funcționale sau previziuni exacte și pentru valori individuale. În silvobiologie, spre exemplu, este foarte greu să se ajungă la ecuații funcționale, tocmai din cauza complexității fenomenelor biologice, unde lanțul deducțiilor nu poate fi finalizat cu același succes ca la studiul anumitor fenomene fizice.

S-a arătat că forma corelației poate fi liniară sau curbilinie. În primul caz ecuația de corelație este ecuația unei linii drepte. În al doilea caz putem întîlni o mulțime de curbe și ecuații corespunzătoare, cum sînt: parabolele de diferite ordine, iperbolele, ecuațiile exponențiale, ecuațiile logaritmice etc.

Stabilirea *tipului funcției de regresie* este o operație importantă, dar și dificilă. Se pornește de la o analiză atentă a tabelei de corelație și a reprezentărilor grafice corespunzătoare care pot fi puse alături de anumite tipuri de curbe întîlnite în geometria analitică. Totodată, deosebit de utile sînt cunoștințele profesionale privind legăturile dintre caracteristicile luate în considerare; ne sprijinim deci și pe studii deductive care de multe ori pot orienta spre alegerea unui tip adecuat de ecuație. De pildă, la compensarea analitică a curbei volumelor într-un arboret echien nu vom folosi niciodată ecuația liniei drepte; de la început ne vom orienta spre ecuația unei parabole de tipul $v = a_o + a_1 d + a_2 d^2$ sau de forma $v = a_o + a_1 d^2$, deoarece se știe că diametrul în formula volumului intervine ridicat la pătrat.

De multe ori sînt necesare mai multe încercări pînă ce se obține un rezultat satisfăcător. În acest scop analiza dispersională poate fi de mare folos, deoarece permite o verificare dacă o anumită formă este mai bună decît alta (R a n c u — T ö v i s s i, 1963).

Scanned with OKEN Scanner

După stabilirea formei adecvate problema se reduce la determinarea coeficienților din ecuația de regresie aleasă.

Ecuațiile de regresie au un caracter empiric, bazîndu-se numai în parte pe cunoașterea explicativă a fenomenelor cercetate. Prin aceasta nu trebuie să se minimalizeze importanța lor științifică și practică. Din contră, acestea sînt deosebit de utile pentru sintetizarea și generalizarea anumitor cunoștințe dobîndite în urma cercetărilor efectuate. Prezentarea rezultatelor cercetării sub această formă *concentrată* se dovedește folositoare nu numai în dendrometrie, unde prin astfel de prezentări se înlocuiesc sute de pagini de valori tabelare, dar și în silvobiologie, studiul lemnului, precum și în multe alte discipline forestiere. În ultimul timp importanța ecuațiilor de regresie a crescut și mai mult, mai ales că prin intermediul lor se pot da soluții interesante și avantajoase pentru automatizarea multor lucrări de calcule biometrice.

Desigur, formulele empirice fiind stabilite în baza unor date experimentale concrete, nu pot fi folosite pentru extrapolări, decît cu foarte mare prudență.

7.2. CITEVA CONSIDERAȚII GENERALE PRIVIND ALEGEREA TIPULUI ECUAȚIEI DE REGRESIE

Nu se pot da reguli generale cu ajutorul cărora să se stabilească tipul cel mai adecvat al ecuației de regresie. Menționăm că formulele empirice pot fi deosebit de variate deoarece la alegerea lor nu sînt posibile deducții teoretice finalizate, importantă prezentînd o singură condiție de bază: obținerea unei concordanțe cît mai bune între valorile observate și cele calculate.

Formulele empirice pot avea un număr diferit de constante, dar cele mai indicate sînt formulele care conțin un număr mai redus de coeficienți; astfel se simplifică atît faza de stabilire a coeficienților, cît și modalitatea de folosire a acestor formule în practică. Dar, atunci cînd formulele sînt destinate unor calcule biometrice de mare serie, cum sînt cele dendometrice frecvent întîlnite în activitatea practică, nu trebuie de mers pe linia simplificării în dauna preciziei; numărul mare de coeficienți sau formulele mai complicate, dacă acestea sînt reclamate de necesitatea ameliorării preciziei, în noile condiții cînd asemenea lucrări de calcule se pot automatiza prin folosirea calculatoarelor electronice, nu mai constituie o piedică în calea folosirii lor în practică.

Frecvent folosit în practica statistico-matematică este polinomul de ordinul n:

$$y = a_0 + a_1 x + a_2 x^2 + \ldots a_n x^n \qquad (7.1)$$

unde a_0, a_1, $a_2 \ldots a_n$ reprezintă coeficienții polinomului (valorile acestor coeficienți se determină pe cale experimentală).

Sînt însă posibile și exprimări mai simple. Pentru alegerea lor prezentăm în continuare cîteva indicații generale.

Scanned with OKEN Scanner

1. Valorile experimentale, pentru care urmează să se aleagă tipul ecuației de regresie, se reprezintă grafic, folosind fie valorile medii $\bar{y}$ calculate pentru categoriile variabilei x, fie valorile individuale negrupate, atunci cînd numărul valorilor observate este redus.

2. În cazul cînd punctele experimentale se așază pe grafic de-a lungul unei linii drepte sau în apropiere de o dreaptă, atunci ne folosim de ecuația:

$$y = a_0 + a_1 x \tag{7.2}$$

Pentru exemplificare prezentăm în fig. 7.1 legătura dintre diametrul de bază măsurat cu coajă și diametrul de bază considerat fără coajă la duglasul verde (*Pseudotsuga menziesii*, Mirbel Franco) cultivat în regiunea Banat. Un alt exemplu (fig. 7.2) se referă la corelația dintre diametrul de bază și lățimea ultimelor 5 inele anuale la un arboret echien de pin.

3. În cazul cînd punctele experimentale nu se ordonează rectiliniar se întocmesc noi grafice, pe hîrtie logaritmică, semilogaritmică, sau folosind alte transformări ale variabilelor x și y. Adesea, punctele așezate pe unul din aceste grafice se ordonează rectiliniar, sau foarte aproape de o linie dreaptă. În acest caz deosebim mai multe situații:

a) Dacă pe hîrtia semilogaritmică punctele experimentale se ordonează de-a lungul unei linii drepte, atunci compensarea analitică se poate realiza după ecuațiile:

$$y = a_1 e^{nx} + a_o \quad \text{(uneori } a_o = 0\text{)} \tag{7.3}$$

$$y = a_0 + a_1 \log x \tag{7.4}$$

În fig. 7.3, pe hîrtie semilogaritmică se prezintă legătura dintre diametrele și mediile înălțimilor măsurate într-un arboret echien de molid. În total au fost măsurate 389 înălțimi (G i u r g i u, 1956). Modul în care

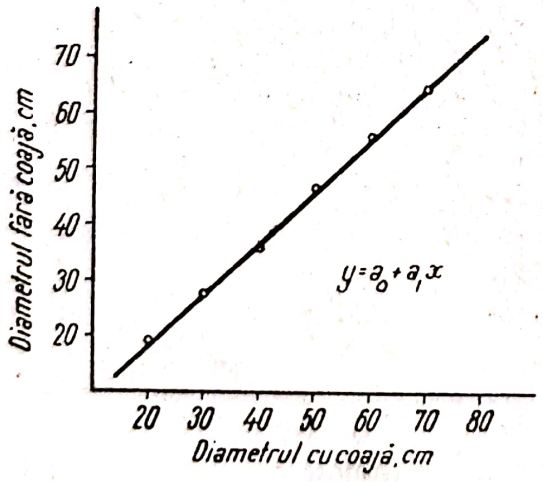


Fig. 7.1. Corelația dintre diametrul de bază cu coajă și diametrul de bază considerat fără coajă la duglasul verde cultivat în Banat

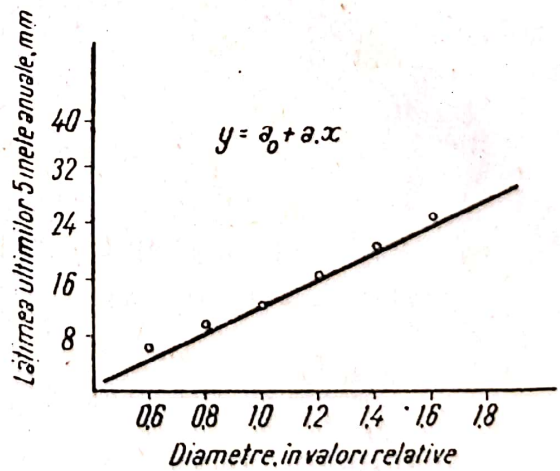


Fig. 7.2. Corelația dintre diametrul de bază și lățimea ultimelor 5 inele anuale (arboret de pin, 30 ani)

Scanned with OKEN Scanner

sînt ordonate punctele experimentale pe grafic, dovedește că în acest caz pentru ajustarea curbei înălțimilor ar putea fi folosită ecuația de tipul (7.4).

b. În cazul ordonării rectiliniare a punctelor experimentale pe *hîrtie logaritmică*, poate fi încercată ecuația:

$$y = a_0 + a_1 x^n \tag{7.5}$$

Uneori $a_0 = 0$.

În fig. 7.4, pe hîrtie logaritmică, se prezintă legătura dintre diametre și volume pentru un arboret echien de molid. Se constată că pentru categoriile de diametre superioare ecuația (7.5) nu corespunde punctelor experimentale.

c) Dacă punctele experimentale se ordonează rectiliniar pe un grafic unde pe axa ox se trec valorile $1:x$, atunci poate corespunde ecuația:

$$y = a_0 + \frac{a_1}{x} \tag{7.6}$$

d) Ecuația

$$y = \frac{1}{a_0 + a_1 x} \tag{7.7}$$

ar putea corespunde dacă punctele experimentale se ordonează rectiliniar pe un grafic unde pe axa oy se trec valorile $1:y$.

e) Ecuația

$$y = \frac{x}{a_0 + a_1 x} \tag{7.8}$$

ar putea corespunde atunci cînd se realizează o linie dreaptă, sau aproape o linie dreaptă, pe un grafic unde pe axa ox trecem valorile $1:x$, iar pe axa oy valorile $1:y$.

Prin astfel de transformări se pot găsi soluții și pentru alte tipuri de ecuații ($y = e^{a_0 + a_1 x}$, $y = a_0 + a_1 x^2$ etc.)

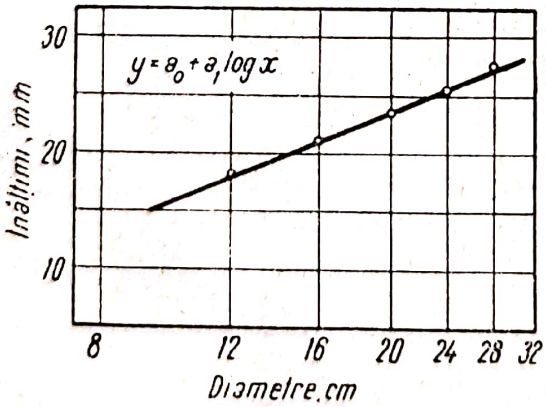


Fig. 7.3. Hîrtie semilogaritmică (corelația dintre diametre și înălțimi la un arboret echien de molid)

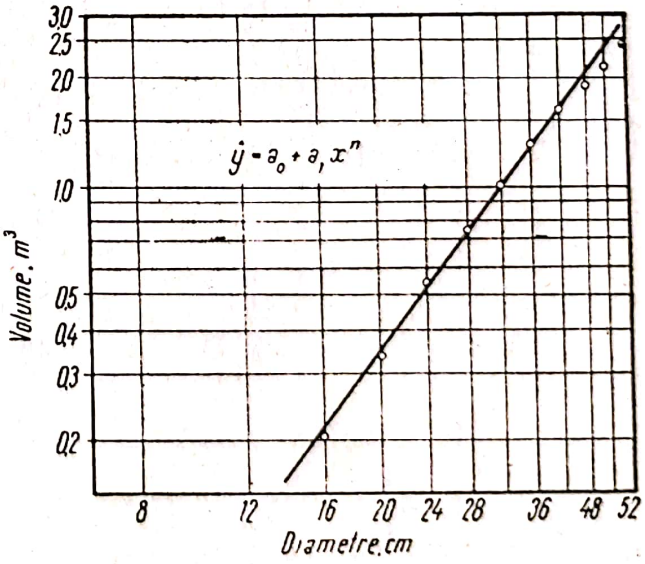


Fig. 7.4. Legătura dintre diametre și volume la un arboret echien de molid

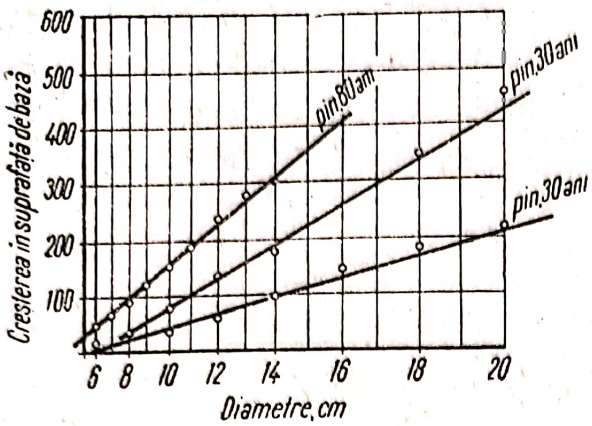


Fig. 7.5. Corelația dintre diametre și creșterea în suprafața de bază la trei arborete echiene (Giurgiu, 1957).

Spre exemplu, în fig. 7.5 se prezintă corelația dintre diametre și creșterea în suprafața de bază la trei arborete echiene, folosind ecuația

$$y = a_0 + a_1 d^2 \qquad (7.9).$$

Să se remarce că pe axa ox s-au trecut valorile d^2.

S-a arătat că în multe situații poate fi folosit polinomul de ordinul n. Pentru a stabili ordinul acestui polinom sînt utile următoarele indicații:

— datele experimentale se transpun pe un grafic;

— din acest grafic, la intervale egale ale variabilei x, se citesc valorile variabilei y (uneori curba obținută se compensează grafic);

— cu aceste date se întocmește tabelul diferențelor succesive;

— se urmărește obținerea unei constanțe aproximative a seriei finale de diferențe; constanța diferențelor reprezintă un indiciu că ordinul polinomului corespunde ordinului diferețelor respective.

Pentru serii tabelate nu mai sînt necesare primele două operații.

Un exemplu concret, privind corelația dintre diametrele și înălțimile relative la arboretele pluriene de fag poate ilustra modul de lucru (tabelul 7.1).

Din tabelul 7.1 rezultă că diferențele de ordinul III sînt destul de apropiate, ceea ce sugerează ideea că o ecuație de forma

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \qquad (7.10)$$

ar putea fi satisfăcătoare.

Lista formulelor posibil de adoptat este, desigur, mult mai mare. Uneori, mai pot fi folosite ecuațiile:

$$y = \frac{1}{a_0 + a_1 x + a_2 x^2} \qquad (7.11)$$

Tabelul 7.1

Stabilirea concordanței dintre valorile experimentale și ecuația de forma $y = a_0 + a_1 x + a_2 x^2 \ldots + a_n x^n$ (corelația dintre diametrele și înălțimile relative la arboretele pluriene de fag)

Diametre x	Înălțimi relative y	Diferențe de ordinul I	Diferențe de ordinul II	Diferențe de ordinul III
16	49			
		19		
24	68		—6	
		13		2
32	81		—4	
		9		3
40	90		—1	
		8		—1
48	98		—2	
		6		—1
56	104		—3	
		3		2
64	107		—1	
		2		1
72	109		0	
		2		0
80	111		0	
		2		
88	113			

$$y = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} \tag{7.12}$$

$$y = \frac{x}{a_0 + a_1 x + a_2 x^2} \text{ sau } y = \frac{x^2}{a_0 + a_1 x + a_2 x^2} \tag{7.13}$$

$$y = ae^{nx + mx^2} \tag{7.14}$$

$$\log y = a_0 + a_1 x + a_2 x^2 \tag{7.15}$$

$$y = a_0 + a_1 x + a_2 \log x \tag{7.17}$$

Majoritatea acestor ecuații pot fi aduse la forme mai simple. Astfel, putem obține o ecuație de forma:

$$y = a_0 + a_1 x + a_2 x^2$$

prin următoarele înlocuiri : în ecuația (7.11) se va introduce $y = 1:z$, în ecuația (7.12). $z = 1:x$, iar în ecuația (7.13) se va introduce variabila $z = x : y$.

În biometrie sînt frecvent folosite funcții susceptibile de o interpretare biologică. Cunoscute sînt:

— curbele de saturație de forma:

$$y = A(1 - e^{-kx}) \tag{7.18 a}$$

— logistica:

$$y = \frac{1}{1 + e^{-(\alpha + \beta x)}} \tag{7.18 b}$$

— funcția de dezvoltare propusă de B a c k m a n:

$$\log y = K_0 + K_1 \log x + K_2 \log^2 x \tag{7.18 c}$$

unde:

x — reprezintă vîrsta; y — creșterea;

A, α, β, K_0, K_1 și K_2 — coeficienți.

Formula lui B a c k m a n prezintă interesante aplicații la studiul creșterii arborilor și arboretelor (P r o d a n, 1961; W e c k, 1955).

Pentru exprimarea matematică a procesului de creștere au fost propuse și alte formule, cum sînt cele prezentate de A s s m a n n (1961), P r o d a n (1961), W e c k (1955). Unele dintre aceste formule rămîn însă din multe puncte de vedere „cămăși de forță" în care silim să intre fenomenul biologic.

7.3. ESTIMAREA COEFICIENȚILOR DIN ECUAȚIILE DE REGRESIE

Pentru estimarea coeficienților din ecuația de regresie aleasă se cunosc mai multe procedee. Foarte simple, dar uneori insuficient de precise, sînt *procedeul punctelor alese* și *procedeul mediilor*. Cea mai sigură este însă *metoda celor mai mici pătrate*.

Scanned with OKEN Scanner

Procedeul punctelor alese se poate folosi pentru toate tipurile de ecuații, dar cele mai bune rezultate se obțin în cazul regresiilor liniare sau pentru ecuațiile care pot fi aduse la forma liniară. Procedeul constă în următoarele:

— Se construiește un grafic, y în funcție de x, alegînd linia cea mai corespunzătoare punctelor experimentale amplasate pe grafic.

— Pe graficul astfel construit se aleg m puncte, unde m reprezintă numărul coeficienților din ecuația aleasă. Se recomandă ca prin alegerea acestor puncte, diferențele dintre abscise să fie aproximativ egale, dar excluzînd extremitățile curbei.

— Cu ajutorul coordonatelor acestor puncte, se formează un sistem de m ecuații cu m necunoscute. Rezolvarea acestui sistem permite determinarea coeficienților căutați.

Pentru a ilustra procedeul descris să considerăm dreapta din fig. 7.1, unde se reprezintă legătura dintre diametrele cu coajă și cele fără coajă la duglasul verde cultivat în regiunea Banat. Ecuația aleasă este de tipul:

$$y = a_0 + a_1 x.$$

Vom alege, deci, două puncte:

$$x_1 = 30 \qquad x_2 = 60$$
$$y_1 = 27{,}5 \qquad y_2 = 54{,}5$$

Formăm un sistem de două ecuații cu două necunoscute:

$$27{,}5 = a_0 + 30\ a_1$$
$$54{,}5 = a_0 + 60\ a_1$$

Rezolvînd acest sistem obținem $a_0 = 0{,}5$, $a_1 = 0{,}9$

Prin urmare

$$\hat{y} = 0{,}9x + 0{,}5 \tag{7.19}$$

În cazul *procedeului mediilor*, se aleg mai multe puncte; uneori pot fi luate în considerare toate perechile de valori observate.

Pentru exemplul din fig. 7.2, luăm în considerare toate cele 6 puncte experimentale :

x:	0,6	0,8	1,0	1,2	1,4	1,6
y:	3	9	12	16	21	25

În baza acestor date, rezultă 6 ecuații inițiale:

$$3 = a_0 + 0{,}6\ a_1$$
$$9 = a_0 + 0{,}8\ a_1$$
$$12 = a_0 + 1{,}0\ a_1$$
$$16 = a_0 + 1{,}2\ a_1$$
$$21 = a_0 + 1{,}4\ a_1$$
$$25 = a_0 + 1{,}6\ a_1$$

Scanned with OKEN Scanner

Formînd două grupe, obținem :

$$24 = 3\,a_0 + 2{,}4\,a_1$$

$$62 = 3\,a_0 + 4{,}2\,a_1$$

De aci rezultă:

$$\hat{y} = -8{,}8 + 21\,x \qquad (7.20\ a)$$

Grupele se pot forma în diferite modalități, însumînd — spre exemplu — ecuațiile pe sărite ($1 + 3 + 5$ și $2 + 4 + 6$). În acest caz obținem ecuația :

$$\hat{y} = -11{,}5 + 23{,}5\,x \qquad (7.20\ b)$$

Numărul ecuațiilor de acest gen poate fi mult mărit. Nu se stabilesc însă mai mult de 3—4 ecuații. Dintre acestea se alege cea mai corespunzătoare, respectiv ecuația pentru care se constată

$$\Sigma\,\Delta_i^2 = min \qquad (7.21)$$

unde Δ_i reprezintă diferențele dintre valorile y calculate ($\hat{y}$) și valorile y experimentale.

Pentru exemplul luat, avem: (tabelul 7.2)

Tabelul 7.2

x	0,6	0,8	1,0	1,2	1,4	1,6	Σ
y	3	9	12	16	21	25	—
$\hat{y}_1$	3,8	8,0	12,2	16,3	20,6	24,8	—
$\hat{y}_2$	2,6	7,2	12,0	16,6	21,3	24,1	—
Δ_1	+0,8	—1,0	+0,2	+0,3	—0,4	—0,2	—
Δ_2	—0,4	—1,7	0	—0,6	+0,3	—0,9	—
Δ_1^2	0,64	1,00	0,04	0,09	0,16	0,04	1,97
Δ_2^2	0,16	2,89	0	0,36	0,09	0,81	4,31

Formula empirică (7.20 a) este mai corespunzătoare decît formula (7.20 b), deoarece $\Sigma\,\Delta_1^2 < \Sigma\,\Delta_2^2$.

7.4. METODA CELOR MAI MICI PĂTRATE

Metoda celor mai mici pătrate conduce nemijlocit la cea mai bună ecuație, respectiv realizează condiția $\Sigma\,\Delta_i^2 = \min$.

În plus, în cazul metodei celor mai mici pătrate, se pot atribui ponderi mai mari sau mai mici anumitor zone din grafic. Singurul dezavantaj însemnat este volumul mare de muncă pe care îl reclamă.

Scanned with OKEN Scanner

Determinarea coeficienților se face cu ajutorul *ecuațiilor normale*. În cazul regresiei liniare, ecuațiile normale sînt:

$$\begin{aligned} n\,a_0 + a_1 \Sigma\, x &= \Sigma\, y \\ a_0 \Sigma\, x + a_1 \Sigma\, x^2 &= \Sigma\, x\, y \end{aligned} \qquad (7{,}22\ a)$$

Pentru parabola de ordinul doi avem:

$$\begin{aligned} n\,a_0 + a_1 \Sigma\, x + a_2 \Sigma\, x^2 &= \Sigma\, y \\ a_0 \Sigma\, x + a_1 \Sigma\, x^2 + a_2 \Sigma\, x^3 &= \Sigma\, x\, y \\ a_0 \Sigma x^2 + a_1 \Sigma\, x^3 + a_2 \Sigma\, x^4 &= \Sigma\, x^2\, y \end{aligned} \qquad (7.22\ b)$$

Prin analogie se poate scrie sistemul de ecuații normale pentru o parabolă de ordinul trei, patru ș.a.

În cazul iperbolei de forma $y = a_0 + a_1 \frac{1}{x}$ scriem:

$$n\,a_0 + a_1 \Sigma \frac{1}{x} = \Sigma\, y$$

$$a_0 \Sigma \frac{1}{x} + a_1 \Sigma \frac{1}{x^2} = \Sigma \frac{1}{x}\, y \qquad (7.23)$$

În mod similar, pentru ecuația de forma:

$$y = a_0 a_1^x \qquad (7.24)$$

sau

$$log\ y = log\ a_0 + x\ log\ a_1$$

avem

$$\begin{aligned} n\ log\ a_0 + log\ a_1 \Sigma\, x &= \Sigma\ log\ y \\ log\ a_0 \Sigma\, x + log\ a_1 \Sigma\, x^2 &= \Sigma\, x\ log\ y \end{aligned} \qquad (7.25)$$

În cazul regresiei liniare de forma:

$$y = a_0 + a_1 x + a_2 z + a_3 u \qquad (7.26)$$

sistemul de ecuații normale va fi:

$$\begin{aligned} n\,a_0 + a_1 \Sigma\, x + a_2 \Sigma\, z + a_3 \Sigma\, u &= \Sigma\, y \\ a_0 \Sigma\, x + a_1 \Sigma\, x^2 + a_2 \Sigma\, x\, z + a_3 \Sigma\, x\, u &= \Sigma\, x\, y \\ a_0 \Sigma\, z + a_1 \Sigma\, x\, z + a_2 \Sigma\, z^2 + a_3 \Sigma\, z\, u &= \Sigma\, z\, y \\ a_0 \Sigma\, u + a_1 \Sigma\, x\, u + a_2 \Sigma\, z\, u + a_3 \Sigma\, u^2 &= \Sigma\, u\, y \end{aligned} \qquad (7.27)$$

Pentru ecuația mai simplă, de forma:

$$y = a_0 + a_1 x + a_2 z \qquad (7.28)$$

Scanned with OKEN Scanner

sistemul se reduce la

$$\begin{aligned} n\,a_0 + a_1 \Sigma x + a_2 \Sigma z &= \Sigma y \\ a_0 \Sigma x + a_1 \Sigma x^2 + a_2 \Sigma x z &= \Sigma x y \\ a_0 \Sigma z + a_1 \Sigma x z + a_2 \Sigma z^2 &= \Sigma z y \end{aligned} \qquad (7.29)$$

În cazul în care numărul perechilor de valori (x, y) este mare, ceea ce frecvent întîlnim în cursul cercetărilor din domeniul biometriei forestiere, calculul coeficienților din ecuațiile de regresie se efectuează în baza tabelului de corelație. În acest caz în sistemul de ecuații normale intervine în plus frecvența n corespunzătoare claselor formate pentru variabila x. Astfel, pentru regresia liniară avem

$$\begin{aligned} a_0 \Sigma n + a_1 \Sigma x n &= \Sigma \bar{y} n \\ a_0 \Sigma x n + a_1 \Sigma x^2 n &= \Sigma x \bar{y} n \end{aligned} \qquad (7.30)$$

În această formulă prin Σn înțelegem numărul total al perechilor de valori (x, y) observate, iar prin $\bar{y}$ se înțelege mediile valorilor y calculate pentru fiecare clasă a variabilei x.

Prin analogie se pot adapta și celelalte sisteme de ecuații normale.

Valorile parametrilor din ecuațiile de regresie alese se determină rezolvînd sistemele de ecuații normale corespunzătoare. În acest scop se pot folosi diferite metode, ca: metoda clasică a eliminării succesive, metoda determinanților, metoda matricială, metoda M o n t e C a r l o, metoda N e m c i n o v etc. Sistemele mai voluminoase se rezolvă cu ajutorul calculatoarelor electronice.

Cînd pentru aceleași valori observate se stabilesc mai multe ecuații de regresie, în final se va alege aceea pentru care se constată cea mai mică valoare a expresiei $\Sigma \Delta_i^2$, unde Δ_i reprezintă diferențele dintre valorile calculate ($\hat{y}$) și valorile $\bar{y}$ experimentale (exemplu, tabelul 7.2). O alegere judicioasă se poate realiza prin intermediul analizei dispersionale (R a n c u-T ö v i s s i, 1963).

E x e m p l u l 1. Într-un arboret echien de molid la 216 arbori s-au măsurat atît diametrele de bază (x) cît și indicii de formă q_2 (y). Rezultatele obținute s-au centralizat într-un tabel de corelație; pentru fiecare categorie de diametre s-a calculat valoarea medie a indicelui de formă. Datele astfel obținute sînt prezentate în tabelul 7.3. Pentru ajustare analitică alegem ecuația liniei drepte.

Desfășurarea calculelor este cît se poate de simplă, însă necesită un volum mare de muncă ce trebuie efectuat cu o deosebită atenție.

Introducînd sumele astfel obținute în sistemul (7.30) rezultă:

$$147\,991 = 216\ a_0 + 6\,492\ a_1$$

$$4\,424\,876 = 6\,492\ a_0 + 3\,206\,800\ a_1$$

Scanned with OKEN Scanner

Tabelul 7.3

Determinarea coeficienților pentru regresia liniară prin metoda celor mai mici pătrate (corelația dintre diametre și indicii de formă la un arboret echien de molid)

x	n	$\bar{y}_{0,...}$	$x\,n$	$x^2\,n$	$\bar{y}\,n$	$x\,\bar{y}\,n$
16	6	724	96	1 536	4 344	69 504
20	29	706	580	11 600	20 474	409 480
24	36	697	864	20 736	25 092	602 208
.						.
.						.
.						.
Total	216	—	6 992	206 800	147 991	4 424 876
	Σn		$\Sigma x n$	$\Sigma x^2 n$	$\Sigma n \bar{y}$	$\Sigma n x \bar{y}$

Rezolvînd sistemul prin procedeul eliminării succesive, obținem:

$$a_0 = 744{,}33\ ; \qquad a_1 = -1{,}97$$

Deci:

$$\hat{y} = 744{,}33 - 1{,}97\ x \tag{7.31}$$

Dacă avem în vedere că $x = d_1$, iar $\bar{y} = 1\,000\ q_2$ * scriem:

$$1\,000\ q_2 = 744{,}33 - 1{,}97\ d_1$$

sau

$$\hat{q}_2 = 0{,}74433 - 0{,}00197\ d_1 \tag{7.32}$$

O comparație între valorile experimentale și cele calculate după formula empirică (7.31) se prezintă în tabelul 7.4 și figura 7.6

Tabelul 7.4

Comparația între valorile experimentale $\bar{y}$ și valorile $\hat{y}$, obținute cu ajutorul ecuației de regresie (7.32)

x	$\bar{y}_{0,...}$	n	$\hat{y}_{0,...}$	$\bar{y} - \hat{y}$	$(\bar{y} - \hat{y})^2$	$(\bar{y} - \hat{y})^2\,n$
16	724	6	713	11	121	726
20	706	29	705	1	1	29
24	697	26	697	0	0	0
.						.
.						.
.						.
Σ	—	216	—	—		18 319

* Pentru simplificare, în tabelul 7.3, valorile indicilor de formă q_2 au fost înmulțite cu 1 000

Scanned with OKEN Scanner

Suma pătratelor diferențelor dintre cele două valori ale variabilei y este relativ mare, mai ales din cauza abaterilor sensibile realizate la extremitatea dreaptă a curbei. De aceea va trebui încercată și altă funcție de regresie (parabola de gradul doi sau o iperbolă), spre a putea compara rezultatele și a alege ecuația cea mai corespunzătoare.

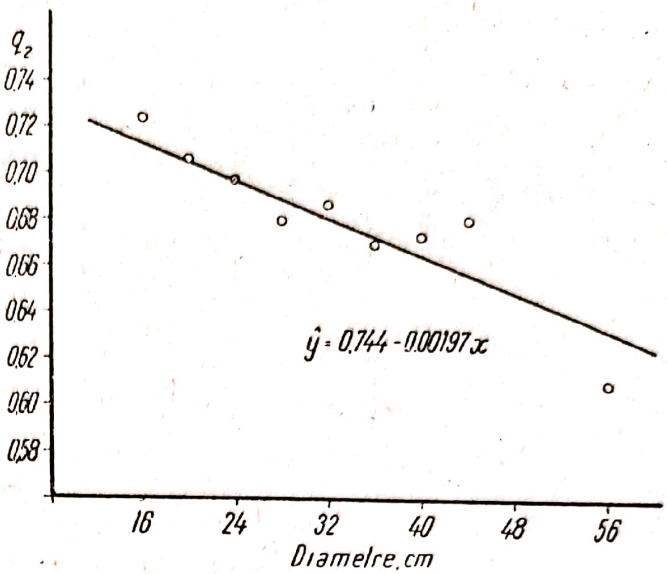


Fig. 7.6. Corelația dintre diametre și indicele pe formă q_2 la un arboret echien de molid.

Exemplul 2. Pentru arboretele echiene Tiurin prezintă date privind corelația dintre diametre și înălțimi, ambele exprimate în valori relative (tabelul 7.5).

Analizînd această corelație ajungem la concluzia că ecuația logaritmică de forma (7.4) ar putea fi corespunzătoare, ceea ce rezultă din figura 7.7 unde pe hîrtia semilogaritmică punctele experimentale s-au ordonat rectiliniar (Giurgiu, 1956).

Desfășurarea calculelor se poate urmări în tabelul 7.6.

Adoptînd această ipoteză, vom folosi sistemul de ecuații normale (7.30), cu deosebirea că în loc de x vom avea $log\ 10\ \chi$.

Introducînd sumele obținute în sistemul (7.30) obținem:

$$987,426 = 979,609\ a_1 + 1\,000\ a_0$$

$$973,469 = 968,7869\ a_1 + 979,609\ a_0$$

Tabelul 7.5

Corelația dintre R_d și R_h la arboretele echiene

R_d	$\overline{R}_h$	n
0,5	0,80	7
0,6	0,85	35
0,7	0,90	95
0,8	0,93	161
0,9	0,97	184
1,0	1,00	181
1,1	1,03	131
1,2	1,06	89
1,3	1,08	63
1,4	1,10	33
1,5	1,12	15
1,6	1,14	5
1,7	1,15	1
Total	—	1 000

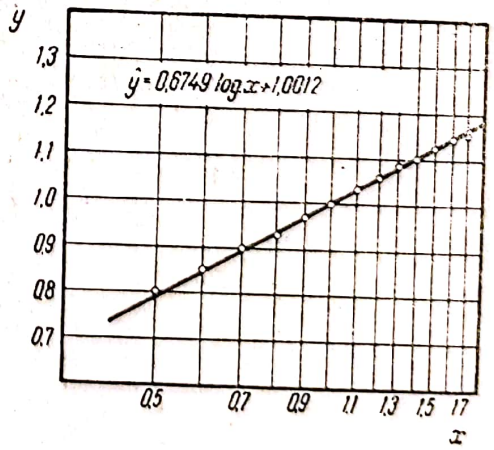


Fig. 7.7. Corelația dintre diametre și înălțimi (exprimate în valori relative) la arborete echiene (pe hîrtie semilogaritmică)

Scanned with OKEN Scanner

Tabelul 7.6

Determinarea coeficienților din ecuația de regresie $y = a_0 + a_1 \log X$ (corelația dintre diametrele și înălțimile exprimate în valori relative (Giurgiu, *1956*).

X	$x = \log(10X)$*	$\bar{y}$	n	nx	nx^2	$n\bar{y}$	$nx\bar{y}$	$\hat{y}$	$\bar{y} - \hat{y}$
0,5	0,699	0,80	7	4,893	3,4202	5,600	3,9144	0,798	—0,002
0,6	0,778	0,85	35	27,230	21,1849	29,750	23,1455	0,851	+0,001
0,7	0,845	0,90	95	80,275	67,8324	85,500	72,2475	0,897	—0,003
.	.	.	.						
.	.	.	.						
.	.	.	.						
			1 000	979,609	968,7869	987,426	973,4690	—	—
		Σ	Σn	Σnx	Σnx^2	Σny	Σnxy	—	—

* Pentru simplificarea calculelor s-a introdus log (10X) în loc de *log* x.

Scanned with OKEN Scanner

de unde rezultă:

$$a_0=0{,}326288$$

$$a_1=0{,}6749$$

așa încît

$$\hat{y}=0{,}326288+0{,}6749\ x$$

Știind că

$$x=log\ (10\ X)$$

scriem

$$\hat{y}=0{,}6749\ log\ X+1{,}0012 \tag{7.32}$$

Din tabelul 7.6 rezultă o foarte bună concordanță între valorile experimentale (coloana 3) și valorile calculate după formula empirică (7.32) (coloana 9). Rezultatele fiind satisfăcătoare, nu mai este necesar să încercăm altă ecuație de regresie. În plus, forma aleasă este simplă, conținînd numai doi coeficienți.

Liniaritatea regresiei este atît de evidentă (fig. 7.7) încît nu mai este necesară verificarea prin criteriul F.

Verificarea ipotezei liniarității regresiei. Referindu-ne din nou la cercetarea legăturii dintre lungimea și greutatea semințelor de salcîm (tab. 6.5) vom stabili ecuația de regresie corespunzătoare. La § 6.2 (tab. 6.6) s-a arătat că legătura dintre aceste două variabile se caracterizează printr-un coeficient de corelație $r=0{,}73$. Pentru aceleași date raportul de corelație este de 0,74 (§ 7.5).

Pentru verificarea ipotezei liniarității regresiei, folosim formula:*

$$F_{y/x}=\frac{(N-n)\ (\eta^2_{y/x}-r^2)}{(n-2)\ (1-\eta^2_{y/x})} \tag{7.33}$$

unde:

N este numărul total de perechi de valori x și y observate;
n — numărul claselor formate pentru variabila x;
$\eta_{y/x}$ — raportul de corelație (§ 7.5).

Valoarea F calculată se compară cu valoarea F tabelară la $f_1=n-2$ și $f_2=N-n$ grade de libertate **.

Pentru exemplul considerat, $N=193$; $n=6$; $r=0{,}73$; $\eta_{y/x}=0{,}74$ (tab. 7.17). Așa încît:

$$F_{y/x}=\frac{(193-6)\ (0{,}74^2-0{,}73^2)}{(6-2)\ (1-0{,}74^2)}=1{,}1$$

---

* Ipoteza liniarității regresiei se poate riguros verifica prin procedeul analizei dispersionale (Loetsch-Haller, 1964; Rancu și Tövissi, 1963). Aici se tratează și problema stabilirii limitelor de încredere pentru regresia liniară, precum și procedeele de comparare a liniilor de regresie.

** Dacă $F_{y/x}$ este subunitar se ia în considerare valoarea inversă ($1/F_{y/x}$), inversîndu-se și numerele gradelor de libertate.

Scanned with OKEN Scanner

Comparăm această valoare cu valoarea tabelară corespunzătoare probabilității $P = 0{,}95$ (nivelul de semnificație 0,05) și în raport cu numărul gradelor de libertate $f_1 = 6-2=4$ și $f_2 = 193-6=187$. În tabelul din anexa 5 găsim $F=2{,}4$.

Întrucît valoarea tabelară F este mai mare decît valoarea F calculată nu avem nici un temei să respingem ipoteza liniarității regresiei variabilei y în raport cu variabila x. Așadar, legătura dintre lungime (x) și greutate (y) la semințele de salcîm studiate se poate exprima cu ajutorul ecuației regresiei liniare.

Stabilirea dreptei de regresie prin intermediul coeficientului de regresie. În cazul cînd coeficientul de corelație a fost calculat, cunoscîndu-se și abaterile medii pătratice (s_x și s_y) ale celor două variabile, precum și mediile lor aritmetice ($\bar{x}$ și $\bar{y}$), ecuațiile de regresie liniară rezultă din relațiile:

$$y - \bar{y} = r \frac{s_y}{s_x} (\bar{x} - \bar{x}) \tag{7.34}$$

$$x - \bar{x} = r \frac{s_x}{s_y} (y - \bar{y}) \tag{7.35}$$

Expresia $r \frac{s_y}{s_x}$ reprezintă coeficientului unghiular al liniei de regresie a variabilei y în raport cu variabila x și poartă denumirea de *coeficient de regresie*. Acesta corespunde coeficientului a_1 din ecuația 7.2, fiind notat frecvent prin simbolul b.

Din tabelul 6.7 avem:

$$s_x = 3{,}99; \quad s_y = 3{,}56$$

$$\bar{x} = 47{,}2; \quad \bar{y} = 21{,}2$$

Așa încît

$$\hat{y} - 21{,}2 = 0{,}73 \frac{3{,}56}{3{,}99} (x - 47{,}2)$$

de unde

$$\hat{y} = 0{,}651\ x - 9{,}5 \tag{7.36}$$

Valorile variabilei x nu se pot deduce algebric din ecuația de regresie (7.36) (acest lucru trebuie în mod deosebit avut în vedere). De aceea, stabilim ecuația de regresie a variabilei x în raport cu variabila y, după formula (7.35):

$$\hat{x} - 47{,}2 = 0{,}73 \frac{3{,}99}{3{,}56} (y - 21{,}2)$$

de unde

$$\hat{x} = 0{,}818\ y + 29{,}9 \tag{7.37}$$

În tabelul 7.7 se prezintă o comparație între valorile experimentale $\bar{y}$ și valorile $\hat{y}$ calculate după ecuația de regresie (7.36) de unde rezultă o bună concordanță între ele.

Scanned with OKEN Scanner

Tabelul 7.7

x	$\bar{y}$	n	$\hat{y}$	$y - \hat{y}$
39	17,0	10	15,8	+1,2
42	17,7	20	17,8	—0,1
45	19,6	55	19,8	—0,2
48	21,4	54	21,7	—0,3
51	23,7	37	23,7	0
54	26,5	12	25,6	+0,9
57	27,8	5	27,6	+0,2
Σ	—	193	—	—

Limitele de încredere pentru dreapta de regresie. O deosebită importanță prezintă stabilirea *limitelor de încredere* ale ecuațiilor de regresie. Pentru regresie liniară acestea se calculează relativ ușor aplicînd formula:

$$s_{\hat{y}} = s_{yx} \sqrt{(1/N) + [(x - \bar{x}^2) / \Sigma [n (x - \bar{x})^2]]} \quad \text{(N — 2 grade de libertate)} \qquad (7.38\ a)$$

unde:

$$s_{yx} = \sqrt{\Sigma [n (\bar{y} - \hat{y})^2] / n - 2} \quad \text{(N — 2 grade de libertate)} \qquad (7.38\ b)$$

În figura 7.9 se prezintă procedeul de calcul al limitelor de încredere ale ecuației de regresie $\hat{y} = -0{,}159 + 0{,}092\ x$, privind corelația dintre diametre și grosimea dublă a cojii (H u s c h, 1964).

Eroarea coeficientului de regresie este dată de formula:

$$s_b = \sqrt{s_{yx}^2} / \sqrt{\Sigma [n (x - \bar{x})^2]} \quad \text{(N—2 grade de libertate)} \qquad (7.38\ c)$$

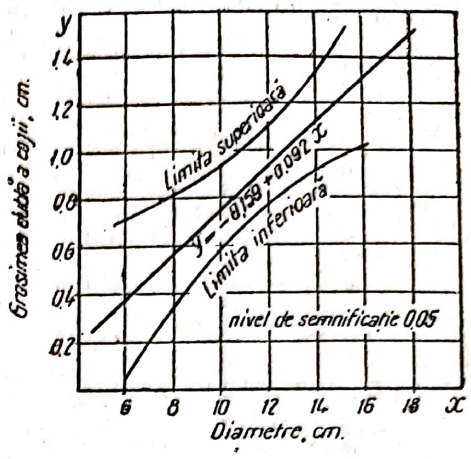
Coeficientul de regresie b este semnificativ dacă b/s este mai mare decît valoarea tabelară a testului t, luată la N—2 grade de libertate (fig. 7.8).

Compararea a doi coeficienți de regresie se efectuează fie după testul $u \left(u = b_1 - b_2 \Big/ \sqrt{s_{b_1}^2 + s_{b_2}^2} \right)$ — atunci cînd volumul selecției este mare, fie după testul t — la selecțiile de volum redus (vezi formulele 4.5 și 4.7, § 4.2). La compararea mai multor coeficienți de regresie sau la compararea mai multor drepte de regresie se aplică analiza dispersională (B a i l e y, 1959; R a n c u — T ö v i s s i, 1963; S n e d e c o r, 1957 ș.a.).

E x e m p l u l 3. La 60 arbori de duglas s-au efectuat măsurători asupra înălțimilor totale h, diametrelor de bază $d_{1,3}$ și asupra diametrelor $d_{0,1}$ luate la 0,1 din înălțimea totală. Pentru fiecare arbore s-a calculat raportul $q = d_{0,1}/d_{1,3}$. Datele astfel obținute au fost grupate în raport cu înălțimea totală, rezultînd datele din tabelul 7.8.

Scanned with OKEN Scanner

Date experimentale			valori calculate				
x	y	n	$\hat{y}$	$(y-\hat{y})n$	$(y-\hat{y})^2 n$	$n(x-\bar{x})$	$n(x-\bar{x})^2$
16.4	1.7	1	1.36	0.34	0.1156	5.14	26.42
13.6	0.9	1	1.10	−0.20	0.0400	2.34	5.48
6.2	0.8	1	0.42	0.38	0.1444	−5.06	25.60
10.2	1.0	1	0.79	0.21	0.0441	−1.06	1.12
12.4	1.2	1	0.99	−0.21	0.0441	1.14	1.30
15.4	1.2	1	1.27	0.07	0.0049	4.14	17.14
10.5	1.0	1	0.81	−0.19	0.0361	−0.76	0.58
10.2	0.5	1	0.79	−0.29	0.0841	−1.06	1.12
11.4	0.6	1	0.90	−0.30	0.0900	0.14	0.02
7.7	0.4	1	0.56	−0.16	0.0256	−3.57	12.74
12.0	0.8	1	0.95	−0.15	0.0225	0.74	0.55
7.3	0.4	1	0.52	−0.12	0.0144	−3.97	15.76
13.1	1.0	1	1.05	−0.05	0.0025	1.89	3.39
146.4	11.5	13	—	—	0.6683	0.00	111.22



de unde: $\bar{x} = 11.26$; $\bar{y} = 0.88$; $b = 0.092$; $S^2_{y/x} = \frac{0.6683}{13-2} = 0.0608$

$\hat{y} = 0.159 + 0.092x$; $S_b = \sqrt{0.0608} / \sqrt{111.22} = 0.024$; $\frac{b}{S_b} = \frac{0.092}{0.024} = 3.8 > 2.2$

$N = 13$

x	$\hat{y}$	$x-\bar{x}$	$(x-\bar{x})^2$	$\frac{(x-\bar{x})^2}{\sum[n(x-\bar{x})^2]}$	$\frac{1}{N} + \frac{(x-\bar{x})^2}{\sum[n(x-\bar{x})^2]}$	$S^2_{\hat{y}}$	$S_{\hat{y}}$	$tS_{\hat{y}}$	Limite de încredere	
									inferioară	superioară
6	0.39	5.26	27.67	0.249	0.326	0.0198	0.147	0.32	0.07	0.71
8	0.58	3.26	10.63	0.096	0.172	0.0105	0.102	0.22	0.36	0.80
10	0.76	1.26	1.59	0.014	0.091	0.0055	0.074	0.16	0.60	0.92
12	0.94	0.74	0.55	0.005	0.082	0.0050	0.071	0.16	0.78	1.10
14	1.13	2.74	7.51	0.068	0.144	0.0088	0.093	0.20	0.93	1.33
16	1.31	4.74	22.47	0.202	0.279	0.0169	0.130	0.03	1.02	1.60

Fig. 7.8. Calculul limitelor de încredere pentru ecuația de regresie liniară (nivel de semnificație 0,05)

Tabelul 7.8

Corelația dintre înălțimea totală și coeficientul q la duglas

Coeficientul q y	Înălțimea totală x						Total
	10	15	20	25	30	35	
1,00	10	3					13
0,98		4	3	1			8
0,96		3	4	4	2		13
0,94			3	4	3		10
0,92				1	3	5	9
0,90					2	5	7
Total	10	10	10	10	10	10	60

În scopul simplificării calculelor, introducem variantele:

$$x' = \frac{x - x_0}{h_x}; \quad y' = \frac{y - y_0}{h_y}$$

Scanned with OKEN Scanner

Alegînd $x_0 = 20$ și $y_0 = 0{,}96$ și cunoscînd că $h_x = 5$ și $h_y = 0{,}02$, obținem tabelul de corelație în raport cu noile variabile (tab. 7. 9). Totodată, în acest tabel se calculează valorile $n_{xy}\ y'$, $\overline{y'}_x$ și $\overline{y}_x$.

Tabelul 7.9

Variabila y'	Variabila x'						n_y
	−2	−1	0	1	2	3	
+2	10	3					13
+1		4	3	1			8
0		3	4	4	2		13
−1			3	4	3		10
−2				1	3	5	9
−3					2	5	7
n_x	10	10	10	10	10	10	60
$n_{xy}\ y'$	20	10	0	−5	−15	−25	
$\overline{y}'_x = \frac{n_{xy}\ y'}{n_x}$	2,0	1,0	0	−0,5	− 1,5	− 2,5	
$\overline{y}_x$	1,00	0,98	0,96	0,95	0,93	0,91	

Pentru coloana a treia obținem: $n_{xy}\ y' = 3 \times 2 + 4 \times 1 + 3 \times 0 = 10$; $\overline{y}_x = 10 : 10 = 1{,}0$; $\overline{y}_x = 3 \times 1{,}00 + 4 \times 0{,}98 + 3 \times 0{,}96 = 0{,}98$.

Presupunînd că regresia studiată este de forma

$$y = a_0 + a_1 x + a_2 x^2 \tag{7.39}$$

pentru variabilele introduse obținem

$$y' = a'_0 + a'_1 x' + a'_2 x'^2$$

sau

$$\frac{\hat{y} - y_0}{h_y} = a'_0 + a'_1 \left(\frac{x - x_0}{h_x}\right) + a'_2 \left(\frac{x - x_0}{h_x}\right)^2 \tag{7.40}$$

Sistemul de ecuații normale va fi:

$$\begin{aligned} a'_0 \sum n_x + a'_1 \sum n_x x' + a'_2 \sum n_x x'^2 &= \sum n_x \overline{y}'_x \\ a'_0 \sum n_x x' + a'_1 \sum n_x x'^2 + a'_2 \sum n_x x'^3 &= \sum n_x x' \overline{y}'_x \\ a'_0 \sum n_x x'^2 + a'_1 \sum n_x x'^3 + a'_2 \sum n_x x'^4 &= \sum n_x x'^2 \overline{y}'_x \end{aligned} \tag{7.41}$$

În scopul rezolvării acestui sistem întocmim tabelele ajutătoare 7.10 și 7.11.

Tabelul 7.10

x'	n_x	$n_x x'$	$n_x x'^2$	$n_x x'^3$	$n_x x'^4$
—2	10	—20	40	—80	160
—1	10	—10	10	—10	10
0	10	0	0	0	0
+1	10	10	10	10	10
+2	10	20	40	80	160
+3	10	30	90	270	810
Σ	60	30	190	270	1 150

Tabelul 7.11

x'	$\bar{y}'_x$	n_x	$n_x \bar{y}'_x$	$n_x \bar{y}' x'$	$n_x \bar{y}'_x x'^2$
—2	2,0	10	20	—40	80
—1	1,0	10	10	—10	10
0	0	10	0	0	0
1	—0,5	10	— 5	— 5	— 5
2	—1,5	10	—15	—30	— 60
3	—2,5	10	—25	—75	—225
Σ	—1,5	60	—15	—160	—200

Introducînd sumele obținute în sistemul (7.41) obținem:

$$60\,a'_0 + 30\,a'_1 + 190\,a'_2 = -15$$

$$30\,a'_0 + 190\,a'_1 + 270\,a'_2 = -160$$

$$190\,a'_0 + 270\,a'_1 + 1\,150\,a'_2 = -200$$

de unde:

$$a'_0 = 0{,}193\,;\ a'_1 = -0{,}870\,;\ a'_2 = -0{,}0025$$

Introducînd în (7.40) rezultă:

$$\frac{\hat{y}_x - 0{,}96}{0{,}02} = 0{,}193 - 0{,}870\left(\frac{x-20}{5}\right) - 0{,}0025\left(\frac{x-20}{5}\right)^2$$

de unde:

$$\hat{y}_x = 0{,}964 - 0{,}017\left(\frac{x-20}{5}\right) - 0{,}00005\left(\frac{x-20}{5}\right)^2 \qquad (7.42)$$

Dînd diferite valori variabilei x obținem valorile ajustate ale variabilei y (tab. 7.12). Se constată o mare apropiere între valorile experimentale $\bar{y}$ și cele calculate $\hat{y}$.

Scanned with OKEN Scanner

Tabelul 7.12

x	$\bar{y}_x$	$\hat{y}_x$	$\bar{y}_x - \hat{y}_x$
10	1,000	0,998	0,002
15	0,980	0,981	—0,001
20	0,960	0,964	—0,004
25	0,950	0,947	0,003
30	0,930	0,930	0
35	0,910	0,913	0,003

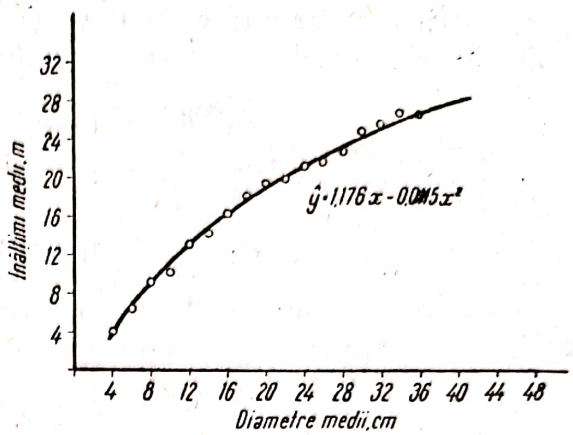


Fig. 7.9. Corelația dintre diametrele medii și înălțimile medii la 297 arborete echiene.

Parabola de ordinul doi a mai fost folosită și pentru ajustarea unei curbe privind variația înălțimilor medii în raport cu diametrele medii la o populație formată din 297 arborete (fig. 7.9).

E x e m p l u l 4. Într-un arboret de molid, la 200 arbori, s-au măsurat: — diametrele de bază x; înălțimile totale z; grosimea dublă a cojii y.

Rezultatele măsurătorilor au fost grupate după cum se vede în tabelul 7.13.

Tabelul 7.13

Corelația dintre grosimea dublă a cojii, diametrul de bază și înălțimea totală, la un arboret de molid

x cm	z m	n	y mm	x cm	z m	n	y mm
12	13	5	7	24	19	10	13
12	15	12	7	24	21	10	11
12	17	13	6	24	23	10	10
16	16	15	10	28	20	7	14
16	18	10	9	28	23	7	13
16	20	15	9	28	26	6	12
20	18	20	12	32	21	4	14
20	20	30	10	32	24	3	14
20	22	20	10	32	26	3	13

Reprezentarea grafică a acestor date arată posibilitatea folosirii ecuației de forma:

$$y = a_0 + a_1 x + a_2 z \qquad (7.28)$$

unde:

y — grosimea cojii;
x — diametrul de bază;
z — lungimea fusului.

Scanned with OKEN Scanner

Pentru determinarea coeficienților a_0, a_1 și a_2 urmează să rezolvăm următorul sistem de ecuații normale:

$$\begin{aligned} a_0 \Sigma n + a_1 \Sigma xn + a_2 \Sigma zn &= \Sigma yn \\ a_0 \Sigma xn + a_1 \Sigma x^2 n + a_2 \Sigma xzn &= \Sigma xyn \\ a_0 \Sigma zn + a_1 \Sigma xzn + a_2 \Sigma z^2 n &= \Sigma yzn \end{aligned} \qquad (7.43)$$

Sumele necesare rezolvării acestui sistem sînt calculate în tabelul 7.14. În baza acestora obținem următorul sistem:

$$\begin{aligned} 200\, a_0 + 4\,000\, a_1 + 3\,907\, a_2 &= 2\,050 \\ 4\,000\, a_0 + 85\,760\, a_1 + 80\,516\, a_2 &= 43\,016 \\ 3\,907\, a_0 + 80\,516\, a_1 + 77\,931\, a_2 &= 40\,684 \end{aligned}$$

Tabelul 7.14

Calculul sumelor necesare rezolvării sistemului (7.43)

x	z	n	y	xn	zn	yn	x^2n	xzn	xyn	z^2n	yzn
12	13	5	7	60	65	35	720	780	420	845	455
12	15	12	7	144	180	84	1 728	2 160	1 008	2 700	1 260
12	17	13	6	156	221	78	1 872	2 652	936	3 757	1 326
.											.
.											.
.											.
Σ	—	200	—	4 000	3 907	2 050	85 760	80 516	43 016	77 931	40 684
	—	Σn	—	Σxn	Σzn	Σyn	Σx^2n	Σxzn	Σxyn	Σz^2n	Σyzn

Rezolvînd acest sistem, obținem: $a_0 = 7{,}085$; $a_1 = 0{,}518$; $a_2 = 0{,}368$. Așa încît

$$\hat{y} = 7{,}085 + 0{,}518\, x - 0{,}368\, z.$$

Această formulă empirică arată că, la același diametru x, grosimea cojii scade pe măsură ce crește înălțimea arborilor.

7.5. RAPORTUL DE CORELAȚIE

Coeficientul de corelație r nu poate exprima intensitatea corelațiilor neliniare. În acest caz gradul de dependență al variabilei y față de variabila x, sau al variabilei x față de y, se măsoară cu ajutorul *raportului de core-*

Scanned with OKEN Scanner

lație $\eta_{y/x}$, respectiv $\eta_{x/y}$. Potrivit definiției, raportul de corelație se calculează după formula:

$$\hat{\eta}_{y/x} = \sqrt{\frac{\hat{\sigma}^2_{y/x}}{\sigma^2_y}} \tag{7.44}$$

$\hat{\sigma}^2_{y/x} = \frac{\Sigma(\hat{y}-\bar{y})^2}{n}$ este dispersia valorilor $\hat{y}$ (calculate după ecuația de regresie adoptată) în jurul mediei $\bar{y}$;

$\sigma^2_y = \frac{\Sigma(y-\bar{y})^2}{n}$ — dispersia valorilor caracteristicii y în jurul mediei $\bar{y}$.

Prin analogie rezultă că

$$\hat{\eta}_{x/y} = \sqrt{\frac{\hat{\sigma}^2_{x/y}}{\sigma^2_x}} \tag{7.45}$$

Pentru calculul raportului de corelație este deci necesar să cunoaștem valorile $\hat{y}$ (respectiv $\hat{x}$) calculate analitic, în care scop se impune stabilirea tipului și a parametrilor ecuației de regresie.

Spre deosebire de coeficientul de corelație, valoarea lui η depinde de alegerea variabilei dependente sau independente. În consecință este necesar să se adauge indicii respectivi spre a se cunoaște care dintre cele două variabile a fost considerată ca variabilă independentă.

Pentru o serie de valori x_i și y_i există o singură valoare a coeficientului de corelație, dar mai multe valori ale raportului de corelație, cîte o valoare separată pentru fiecare ecuație de corelație adoptată. De aceea, valoarea raportului de corelație este legată de o anumită ecuație de regresie.

Mai puțin riguros raportul de corelație se poate calcula și fără a cunoaște forma ecuației de regresie pe baza *curbei mediilor*, aplicînd formula

$$\eta_{\bar{y}/x} = \sqrt{\frac{\sigma^2_{\bar{y}x}}{\sigma^2_y}} \tag{7.46}$$

unde:

$\sigma^2_{\bar{y}x}$ reprezintă dispersia valorilor medii pe grupe ale caracteristicii x în jurul mediei generale $\bar{y}$.

Prin analogie se dovedește că

$$\eta_{\bar{x}/y} = \sqrt{\frac{\sigma^2_{\bar{x}y}}{\sigma^2_x}} \tag{7.47}$$

Aceste două formule se pot folosi pentru o primă estimație a raportului de corelație, atunci cînd cercetătorul nu intenționează să stabilească tipul și parametrii ecuației de regresie respectivă. Cînd ecuația de regresie este determinată, ceea ce se recomandă, trebuie folosite formulele (7.44) și (7.45).

Scanned with OKEN Scanner

De obicei, valorile reale ale dispersiilor σ^2 nu se cunosc; de aceea se înlocuiesc cu estimațiile lor corespunzătoare s^2, obținute pe cale experimentală.

Raportul de corelație are valori cuprinse între 0 și +1. Dacă între caracteristicile x și y există o legătură funcțională, raportul de corelație este egal cu 1. Pentru valori ale raportului de corelație apropiate de 0 se trage concluzia că legătura este slabă sau inexistentă. Cu cît η are valori mai apropiate de 1, cu atît corelația este mai puternică. În cazul unei corelații strînse $\eta_{y/x}$ este apropiat de $\eta_{x/y}$. Diferența $\eta^2 - r^2$ reflectă curbiliniaritatea regresiei.

Calculul raportului de corelație. S-a arătat că raportul de corelație se calculează numai după ce se stabilește tipul ecuației de regresie, precum și parametrii acestei ecuații.

Conform definiției, raportul de corelație se calculează după formula (7.46) unde dispersiile σ^2 sînt înlocuite prin dispersiile s^2:

$$\eta_{y/x}^{\wedge} = \sqrt{\frac{s^2_{\wedge}}{s^2_y}} \qquad (7.48)$$

Dispersia s_y^2 se calculează după formula cunoscută:

$$s_y^2 = \frac{1}{N}\sum n_y (y - \bar{y})^2$$

sau

$$s_y^2 = h_y^2 \left[\frac{\Sigma y'^2 n_y}{\Sigma n_y} - \left(\frac{\Sigma y' n_y}{\Sigma n_y} \right)^2 \right] \qquad (7.49)$$

unde:

y' reprezintă valorile simplificate ale variabilei y.

Această dispersie se poate calcula și după procedeul sumelor, descris la § 2.5.

Dispersia $s^2_{\wedge}$, reprezentînd împrăștierea valorilor $\hat{y}$ (valorile determinate după ecuația de regresie) în jurul mediei $\bar{y}$, se calculează folosind formula:

$$s^2_{\wedge} = \frac{1}{N}\sum n_x (\hat{y} - \bar{y})^2 \qquad (7.50)$$

Tabelul 7.15

y'	n_y	$y'n_y$	$y'^2 n_y$
2	13	26	52
1	8	8	8
0	13	0	0
—1	10	—10	10
—2	9	—18	36
—3	7	—21	63
Σ	60	—15	169

Pentru ilustrare folosim datele exemplului 3 de la § 7.4.

Pentru determinarea raportului de corelație avem nevoie de dispersiile $s^2_{\wedge}$ și s^2_y. În acest scop întocmim tabelele 7.15 și 7.16.

Tabelul 7.16

x	n_x	$\hat{y}_x$	$\hat{y}_x - \bar{y}$	$(\hat{y}_x - \bar{y})^2$	$(\hat{y}_x - \bar{y})^2 n_x$
—2	10	0,998	0,043	0,00185	0,0185
—1	10	0,981	0,026	0,00068	0,0068
0	10	0,964	0,009	0,00008	0,0008
1	10	0,947	—0,008	0,00006	0,0006
2	10	0,930	—0,015	0,00023	0,0023
3	10	0,913	—0,042	0,00176	0,0176
Σ	60	—			0,0466

$\bar{y} = 0{,}955$

După datele din tabelul 7.15 obținem:

$$\bar{y} = y_0 + h_y - \frac{\Sigma y' n_y}{\Sigma n_y} = 0{,}960 - 0{,}02 \cdot \frac{15}{60} = 0{,}955$$

$$s_y^2 = h_y^2 \left[\frac{\Sigma y'^2 n_y}{\Sigma n_y} - \left(\frac{\Sigma y' n_y}{\Sigma n_y} \right)^2 \right] = 0{,}02^2 \left[\frac{169}{60} \left(\frac{-15}{60} \right)^2 \right] = 0{,}001103$$

Aplicînd formula (7.50), avem

$$s^2_{\hat{y}x} = \frac{1}{60} 0{,}0466 = 0{,}000775.$$

În sfîrșit

$$\eta_{\hat{y}/x} = \sqrt{\frac{0{,}000775}{0{,}001103}} = 0{,}835$$

Constatăm, deci, o corelație destul de strînsă între cele două caracteristici studiate, în raport cu ecuația de regresie stabilită (7.42).

Raportul de corelație $\eta_{\hat{y}/x}$ are altă valoare decît 0,835.

*

Dacă nu cunoaștem tipul și coeficienții ecuației de regresie, putem folosi raportul de corelație calculat în baza *curbei mediilor*, așa cum s-a arătat mai sus. În acest scop se aplică formula (7.46). Pentru exemplul considerat avem (vezi tab. 7.9). :

$$s^2_{\bar{y}x} = \frac{1}{60} \Big[10 \Big(1{,}000 - 0{,}955 \Big)^2 + 10 \Big(0{,}980 - 0{,}955 \Big)^2 + \ldots$$

$$\ldots + 10 \Big(0{,}910 - 0{,}955 \Big)^2 \Big] = 0{,}00089$$

$$s_y^2 = 0{,}001103$$

Așa încît

$$\eta_{\bar{y}/x} = \sqrt{\frac{0{,}00089}{0{,}001103}} = 0{,}90$$

Scanned with OKEN Scanner

Raportul de corelație calculat în baza curbei mediilor este mai mare decît cel calculat în raport cu ecuația de regresie stabilită.

Pentru corelația dintre lungimea și greutatea semințelor de salcîm (tab. 6.4) avem:

$$s_y^2 = 12{,}7;\ s_{\bar{y}x}^2 = \frac{1}{193}\Big[10\Big(17{,}0 - 21{,}2\Big)^2 + \\ + 20\Big(17{,}7 - 21{,}2\Big)^2 + \ldots + 5\Big(27{,}8 - 21{,}2\Big)^2\Big] = 6{,}95$$

unde 17,0; 17,7...27,8 reprezintă mediile variabilei y pe clase ale variabilei x.

Așa încît

$$\eta_{\bar{y}/x} = \sqrt{\frac{6{,}95}{12{,}7}} = 0{,}74$$

Un rezultat apropiat s-a obținut și prin *procedeul sumelor*, prezentat în tabelul 7.17. Ultimul procedeu prezintă însemnate avantaje în privința comodității de calcul.

Tabelul 7.17

x / y	39		42		45		48		51		54		57		TOTAL		
	n	(1)	n	(1)	n	(1)	n	(1)	n	(1)	n	(1)	n	(1)	n	(1)	(2)
28									6	6	6	6	3	3	15	15	15
26							2	2	9	15	3	9	2	5	16	31	46
23	1	1			6	6	25	27	10	25	2	11		5	44	75	
20	1	–			37		24		11		1				82		
17	5	8	12	12	10	12	3	3	1	1					28	36	
14	3	3	3	3	2	2									8	8	8
N	10				55		54		37		12		5	$N_t = 193$			
a	1		0		6		29		46		26		13		121		61
b	11		15		14		3		1						44		8
d	-10		-15		-8		26		45		26		13		$d_t = 77$		
d^2	100		225		64		676		2025		676		196		$d_t^2 = 5929$		
$\frac{d^2}{N}$	10		11,2		1,2		12,3		54,8		56,2		39,2		$\frac{d_t^2}{N_t}$ 30,7		
															S 165		69

$A = \sum \frac{d^2}{N} = 10 + 11{,}2 + 1{,}2 + 12{,}3 + 54{,}8 + 56{,}2 + 39{,}2 = 184{,}8$

$B = \frac{d_t^2}{N_t} = 30{,}7$

$C = S_1 + 2S_2 = 165 + 2 \cdot 69 = 303$

$$\eta_{\bar{y}/x} = \sqrt{\frac{A - B}{C - B}} = \sqrt{\frac{184{,}8 - 30{,}7}{303 - 30{,}7}} = 0{,}74$$

VIII

METODA SELECTIVĂ ȘI IMPORTANȚA EI PENTRU SILVICULTURĂ

8. 1. INTRODUCERE

Caracterizarea numerică a fenomenelor mai ales în silvicultură, rareori se poate efectua pornind de la înregistrări totale; de cele mai multe ori, în acest scop se folosește *metoda selectivă*, care constă în estimarea valorilor medii ale colectivității pornind de la examinarea unui *eșantion* (o anumită parte extrasă din colectivitate). În acest caz statistica matematică tinde ca pornind de la cercetarea eșantionului să cunoască întreaga colectivitate. Formulată în termeni matematici, metoda constă în determinarea indicatorilor statistici ai colectivității generale cu ajutorul variabilelor x_i ($i = 1, 2, 3, \ldots, n$) referitoare la n elemente (unități) extrase dintr-o colectivitate generală formată din N elemente ($n < N$). Numărul elementelor din selecție reprezintă *volumul selecției*.

În activitatea practică, silvicultorul s-a folosit frecvent încă din secolele trecute de procedeul sondajelor, cu toate că această practică nu a avut o fundamentare științifică. Astfel, la inventarierea fondului de producție, pentru evitarea inventarierilor integrale, s-a folosit de procedeul suprafețelor de probă, încercînd ca prin probă să se tragă concluzii pentru întreg. În ultimul timp, metoda selectivă a devenit un instrument de lucru de cea mai mare utilitate pentru silvicultor, ceea ce se explică prin:

1) costul redus al metodei în comparație cu înregistrările integrale;
2) imposibilitatea practică de a efectua înregistrări totale pentru multe fenomene și caracteristici;
3) operativitatea cu care se obțin rezultatele;
4) eficacitatea mai mare a metodei;
5) sporirea calității înregistrărilor, deoarece numărul redus de unități observate permite o cercetare atentă și multilaterală a lor.

În domeniul silviculturii, metoda selectivă a fost aplicată pentru prima dată de danezul O p p e r m a n n (1913) și apoi de finlandezul C a j a n u s (1914). Această metodă a luat o deosebită amploare în țările din nordul Europei, în special Suedia (N ä s l u n d, 1930) și Finlanda (I l v e s s a l o,

1923) mai ales la inventarierea fondului de producție pe suprafețe mari (R i c h t e r, 1958). În prezent, metoda selectivă se aplică în majoritatea țărilor, constatîndu-se o creștere a interesului față de această metodă, pe măsură ce se trece la o silvicultură din ce în ce mai intensivă. Într-adevăr, în condițiile unei silviculturi intensive, se impune o înregistrare și o cunoaștere din ce în ce mai amplă a pădurii și a efectului intervențiilor silviculturale asupra stării fondului forestier privit din toate punctele de vedere, ceea ce — desigur — nu se poate realiza decît prin metode bazate pe eșantioane.

Bineînțeles, și în trecut majoritatea cercetărilor din domeniul silviculturii au avut la bază observații și măsurători efectuate prin sondaje. În ultimul timp însă, aceste cercetări au trecut într-o nouă fază, caracterizată prin folosirea multilaterală a metodelor statisticii matematice la organizarea cercetărilor, precum și la interpretarea rezultatelor obținute, arătînd gradul de nesiguranță a concluziilor. În activitatea practică, metoda selectivă va trebui să se aplice nu numai la inventarierea fondului de producție, dar și la controlul calității anumitor produse (controlul calității materialului de împădurire, verificarea calității anumitor sortimente, studiul defectelor lemnului, controlul calității împăduririlor, controlul eficacității măsurilor de combatere a dăunătorilor etc.), precum și la cercetarea utilizării timpului de lucru al muncitorilor, la studiul îndeplinirii normelor de lucru, la fundamentarea normelor de consum, la cercetarea influenței anumitor factori asupra nivelului și dinamicii productivității muncii, la cercetarea influenței înzestrării tehnice a muncii etc.

Metoda selectivă poate aduce foloase importante la inventarierea produselor accesorii ale pădurilor, în special la inventarierea resurselor micologice din fondul forestier, precum și la inventarierea efectivului de vînat.

Metoda selectivă devine unul din factorii hotărîtori în organizarea producției forestiere prin faptul că, prin cheltuieli reduse și într-un timp scurt, furnizează informații deosebit de utile pentru cunoașterea procesului de producție și a factorilor care îl influențează, ceea ce permite luarea de măsuri în scopul obținerii unei producții cantitativ și calitativ superioare.

Principiul selecției se bazează în totalitate pe faptul că eșantionul la care se aplică metodele statistico-matematice de prelucrare este extras la *întîmplare* din colectivitatea totală și că orice alt eșantion de aceeași mărime are aceeași șansă să fie ales. Foarte important de subliniat este faptul că numai pentru selecțiile întîmplătoare este posibil să se determine anumite limite ale erorilor comise; numai astfel se poate aprecia o măsură a gradului de siguranță pentru rezultatele obținute din selecție. Condiția esențială a cercetării statistice constă în asigurarea *reprezentativității* părții selecționate, în care scop se impune ca unitățile extrase să aibă o egală șansă de a face parte din colectivitatea de selecție. În același timp se impune ca unitățile din care este formată colectivitatea generală să fie cît mai *omogene* din punct de vedere al caracteristicilor luate în considerare. O altă condiție se referă la *volumul selecției* care trebuie să fie suficient de mare.

Reprezentativitatea perfectă se asigură atunci cînd structura colectivității de selecție se apropie de structura colectivității generale. Dar o asemenea reprezentativitate bună este greu de realizat sau se realizează

Scanned with OKEN Scanner

numai în cazuri izolate. Uneori nu se poate asigura tuturor unităţilor din colectivitatea generală aceeaşi şansă să ajungă în colectivitatea de selecţie. Din aceste cauze *apar erori de reprezentativitate* care pot fi sistematice sau întîmplătoare. Erorile de reprezentativitate sistematice apar ori de cîte ori nu se asigură tuturor unităţilor din colectivitatea generală aceeaşi probabilitate să facă parte din colectivitatea de selecţie. Asemenea erori pot să apară frecvent în lucrările de inventariere parţială a fondului forestier cînd suprafeţele de probă se amplasează defectuos prin concentrarea lor în partea inferioară a versantului sau în anumite porţiuni ale arboretului cu panta mai mică. Este evident că în aceste cazuri selecţia nu va mai fi reprezentativă, deoarece în partea inferioară a versantului sau pe porţiunile cu panta mai mică volumul la hectar va fi mai mare decît în arboretul întreg. Asemenea erori pot să apară şi atunci cînd regenerarea arboretelor se studiază pe suprafeţe de probă situate la marginea masivului.

Erorile de reprezentativitate întîmplătoare, datorite diferenţelor de structură dintre cele două colectivităţi, spre deosebire de primele, practic nu pot fi evitate decît în cazul unei inventarieri integrale. Marele avantaj al metodei selective constă însă în aceea că mărimea acestor erori poate fi estimată cu anticipaţie, în funcţie de:

— volumul selecţiei;

— variabilitatea unităţilor supuse cercetării selective, respectiv de mărimea coeficientului de variaţie;

— metoda de selecţie folosită;

— procedeele de prelucrare a valorilor observate.

Desigur, reducerea erorilor de reprezentativitate se poate realiza prin majorarea volumului selecţiei, ştiind că există o probabilitate mare, mult apropiată de 1, ca în cazul unui volum suficient de mare de unităţi observate, parametrii colectivităţii de selecţie să fie foarte apropiaţi de parametrii colectivităţii generale. După cum se va vedea, eroarea de reprezentativitate este invers proporţională cu rădăcina pătrată din numărul unităţilor incluse în colectivitatea de selecţie.

Dar majorarea volumului selecţiei este legată de sporirea cheltuielilor; de aceea statistica matematică a elaborat procedee potrivite de selecţie, prin care se pot reduce erorile de reprezentativitate întîmplătoare. Astfel de procedee au la bază ideea *stratificării*, indicînd gruparea unităţilor din colectivitatea generală în aşa fel încît în fiecare strat (grupă) să fie incluse unităţi statistice cît mai omogene. Procedeul stratificării are o însemnătate remarcabilă pentru majoritatea inventarierilor parţiale efectuate în domeniul nostru de activitate, mai ales la inventarierea fondului forestier privit din toate punctele de vedere.

Un alt aspect care trebuie multilateral studiat este cel referitor la rentabilitatea inventarierilor. Se ştie că cheltuielile raportate la unitatea de observare sînt mai mari în cazul selecţiei decît în cazul inventarierilor (înregistrărilor) integrale. În plus mai trebuie avut în vedere unele dificultăţi implicate de organizarea unei inventarieri parţiale. De aceea, selecţiile pentru colectivităţile generale de volum mic sînt uneori nerentabile. Dacă selecţia nu reprezintă decît o fracţiune redusă a colectivităţii generale, pe ansamblu cheltuielile sînt foarte mici în cazul selecţiei în comparaţie

Scanned with OKEN Scanner

cu inventarierea totală. De pildă, inventarierile parțiale ale fondului de producție sînt foarte rentabile în cazul cînd suprafața de inventariat este mare. Din contra, inventarierile parțiale pentru suprafețele reduse, de cel mult 2 — 3 ha, sînt de regulă neindicate, întrucît cheltuielile necesitate de organizarea și efectuarea selecției sînt apropiate sau depășesc cheltuielile unei inventarieri integrale.

După modul în care se efectuează selecția, deosebim mai multe procedee :

8. 2. SELECȚIA INTIMPLĂTOARE NEREPETATĂ

Procedeul reclamă o alegere la întîmplare a colectivității de selecție. Elementul, o dată cercetat, nu se mai înapoiază în colectivitate. Dar această alegere nu înseamnă o luare a sondajului în mod arbitrar. La formarea colectivității de selecție, se urmărește eliminarea subiectivismului și a oricăror manifestări tendențioase. În acest scop se pot folosi regulile unei scheme de loterie, adoptînd practica bilețelelor numerotate și perfect amestecate. O alegere întîmplătoare riguroasă se efectuează și cu ajutorul numerelor întîmplătoare (anexa 9) (S n e d e c o r, 1957; P r o d a n, 1961).

Precizia probabilistică* a selecției nerepetate este dată de formulele:

— pentru caracteristici nealternative

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \tag{8.1}$$

— pentru caracteristici alternative

$$s_p = \sqrt{\frac{p(1-p)}{n}} \sqrt{\frac{N-n}{N-1}} \tag{8.2}$$

unde:

s este abaterea medie practică;
N — volumul colectivității generale;
p — frecvența relativă a evenimentului A.

Pentru colectivități infinite sau atunci cînd $n/N < 0{,}1$, raportul $\frac{N-n}{N-1}$ se apropie de 1, astfel încît în asemenea situații formulele (8.1) și (8.2) devin:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{8.3}$$

$$s_p = \sqrt{\frac{p(1-p)}{n}} \tag{8.4}$$

*) Noțiunea de precizie se referă la erorile de reprezentativitate întîmplătoare, respectiv la variația acestor erori în limita unui interval stabilit cu o probabilitate cunoscută.

Scanned with OKEN Scanner

În aceste formule dispersia generală σ^2 a fost înlocuită cu dispersia de selecție s^2, calculată pe baza datelor selecției, deoarece în cercetările practice întîlnim foarte rar cazuri cînd σ^2 este cunoscută.

Se demonstrează că media de selecție $\bar{x}$ este o variabilă aleatoare și de aceea la alt sondaj întîmplător ea poate lua alte valori, variind însă în anumite limite care cu o probabilitate P nu vor fi depășite. Aceste limite sînt:

$$\bar{x} \pm u\, s_{\bar{x}} \text{ sau } \bar{x} \pm t s_{\bar{x}}$$

unde:

— u sau t se determină în funcție de probabilitatea de depășire admisă, după cum urmează:

— u după anexa 3, în cazul unui număr mare de măsurători;
— t după anexa 4, după selecții de volum mic ($n < 30$).

Produsul $us_{\bar{x}}$ sau $ts_{\bar{x}}$ reprezintă eroarea limită a mediei de selecție, arătînd că mediile de selecție pot lua valori care nu vor depăși media colectivității generale cu mai mult de $\pm us_{\bar{x}}$ sau $\pm ts_{\bar{x}}$; această afirmație este valabilă pentru probabilitățile corespunzătoare lui u sau t alese (vezi § 4.1).

Exemplul 1. Pentru molidul din țara noastră la $n = 947$ arbori aleși la întîmplare s-a determinat mărimea indicelui de formă q_2, rezultînd următoarele valori ale parametrilor distribuției (tab. 3.7):

$$\bar{x} = 0{,}666$$

$$s = 0{,}58.$$

În acest caz, N se apropie de ordinul valorilor infinite, așa încît vom aplica formula (8.3):

$$s_{\bar{x}} = \frac{0{,}058}{\sqrt{947}} = 0{,}0019.$$

Pentru $P = 95\%$ (nivelul de semnificație 5%) obținem $u = 1{,}96$ (anexa 3).

Prin urmare, intervalul de încredere va fi

$$\bar{x} \pm 1{,}96\, s_{\bar{x}} = 0{,}666 \pm 1{,}96 \times 0{,}0019 = 0{,}666 \pm 0{,}004.$$

Aceasta înseamnă că mărimea medie a indicelui de formă la molidul din țara noastră nu va fi mai mică de 0,662 și nici mai mare de 0,670; evident, această afirmație corespunde unei siguranțe probabilistice de 0,95, potrivit distribuției normale.

Pentru siguranța probabilistică de 0,99 limitele intervalului de încredere vor fi: $0{,}66 \pm 2{,}58 \times 0{,}0019 = 0{,}666 \pm 0{,}005$.

Scanned with OKEN Scanner

Problema poate fi pusă și invers: care este probabilitatea ca media de selecție să se găsească într-un anumit interval, determinat de abaterea $\Delta_{\bar{x}}$ față de media colectivității totale.

Potrivit celor cunoscute, această probabilitate este dată de expresia:

$$P = 2\,\Phi(u) = \Phi\left(\frac{\Delta_{\bar{x}}}{s_{\bar{x}}}\right) \tag{8.5}$$

unde valoarea funcției $2\Phi(u)$ se determină după anexa 3.

Revenind la exemplul considerat, să determinăm probabilitatea ca indicele de formă al molidului din Republica Socialistă România să nu depășească media generală cu mai mult de 0,006.

$$u = \frac{\Delta_{\bar{x}}}{s_{\bar{x}}} = \frac{0{,}006}{0{,}0019} = 3{,}2.$$

Valoarea probabilității corespunzătoare acestui u este de 99,86% (anexa 3). Prin urmare, cu probabilitatea de 99,9% (aproape certitudine totală) putem afirma că abaterea mediei de selecție față de media colectivității generale nu va depăși $\Delta_{\bar{x}} = \pm 0{,}006$.

Exemplul 2. Efectuînd inventarieri parțiale într-un arboret plurien de fag, în proporție de 2%, prin amplasarea la întîmplare a 27 suprafețe de probă circulare de 500 m², a rezultat pentru media de selecție valoarea de 27,1 m³, abaterea medie pătratică calculată în baza datelor de selecție fiind de 9,2 m³. De aici rezultă:

$$s_{\bar{x}} = \frac{9{,}2}{\sqrt{27}} = 1{,}8 \text{ m}^3 \text{ sau } 6{,}7\%$$

Avînd în vedere că volumul selecției este mic, pentru stabilirea intervalului de încredere, se utilizează distribuția t. În cazul unei probabilități de transgresiune de 5%, la $f = 27-1 = 26$ grade de libertate, rezultă pentru t valoarea de 2,06 (anexa 4).

Așa încît, limitele intervalului de încredere sînt :

$$27{,}1 \pm 2{,}06 \times 1{,}8 = 27{,}1 \pm 3{,}7.$$

Pentru colectivitatea generală obținem: 3 658 ± 500.

Cu alte cuvinte, fondul de producție al întregului arboret estimat prin inventarierea parțială efectuată nu va fi mai mic de 3 158 m³ și nici mai mare de 4 158 m³ și acest lucru se afirmă cu o probabilitate de 95%.

Pentru o probabilitate de 90% limitele sînt cu mult mai restrînse (t = = 1,71, vezi anexa 4):

$$27{,}1 \pm 1{,}71 \times 1{,}8 = 27{,}1 \pm 3{,}1$$

Exemplul 3. Într-un arboret exploatabil de molid s-au ales la întîmplare 1511 arbori, dintre care 53 au îndeplinit condițiile calitative cerute pentru lemnul de rezonanță. Se cere să se estimeze proporția arborilor

Scanned with OKEN Scanner

de rezonanță cu arătarea limitelor de încredere pentru o probabilitate de 95% (nivel de semnificație 5%)

$$p = \frac{53}{1\,511} = 0{,}035; \qquad q = 1 - p = 0{,}965$$

$$s_p = \sqrt{\frac{0{,}035(1 - 0{,}035)}{1\,511}} = 0{,}0047.$$

Pentru probabilitatea de 95% corespunde $u = 1{,}96$. Așa încît obținem următorul interval de încredere pentru proporția molidului de rezonanță:

$$0{,}035 \pm 1{,}96 \times 0{,}0047 \text{ sau } 0{,}035 \pm 0{,}009.$$

Prin urmare, proporția molidului de rezonanță în arboretul considerat poate lua valori între 0,026 și 0,044; acest lucru se afirmă cu o probabilitate de 95%.

Una din cele mai importante probleme ce se pot rezolva în cadrul metodei selective este *stabilirea volumului selecției*. În acest scop se ține seama de: precizia pe care trebuie să o îndeplinească rezultatele sondajului, respectiv de eroarea admisă Δ_x; probabilitatea cu care urmează a se garanta acest rezultat; cheltuielile necesare efectuării sondajului respectiv.

Pornind de la formulele (8.1) și (8.2) stabilim că volumul selecției se poate determina după cum urmează:

— pentru caracteristici nealternative

$$n = \frac{u^2 s^2 N}{N \Delta_{\bar{x}}^2 + u^2 s^2} \tag{8.6}$$

sau, pentru colectivități infinite (N foarte mare)

$$n = \frac{u^2 s^2}{\Delta_{\bar{x}}^2} \tag{8.7}$$

— pentru caracteristici alternative

$$n = \frac{u^2 \cdot p\,(1 - p)\,N}{N\Delta_p^2 + u^2 p\,(1 - p)} \tag{8.8}$$

sau, pentru colectivități infinite (N foarte mare)

$$n = \frac{u^2 \cdot p(1 - p)}{\Delta_p^2} \tag{8.9}$$

Este foarte ușor de observat că, dacă se urmărește dublarea preciziei, celelalte condiții rămînînd constante, va trebui să mărim volumul selecției de 4 ori. Pentru o triplare a preciziei va trebui să majorăm volumul selecției de 9 ori.

Scanned with OKEN Scanner

Uneori este mai comod să se folosească coeficientul de variație $s\%$ în locul abaterii medii pătratice s, iar în locul lui $\Delta_{\bar{x}}$ se ia eroarea în procente. În această situație formulele (8.6) și (8.7) devin:

$$n = \frac{u^2 s_{\%}^2 N}{N\Delta_{\%}^2 + u^2 s_{\%}^2} \tag{8.10}$$

$$n = \frac{u^2 s_{\%}^2}{\Delta_{\%}^2} \tag{8.11}$$

Precizia pe care trebuie să o îndeplinească rezultatele selecției se stabilește de la caz la caz în funcție de anumite cerințe obiective. Spre exemplu, la inventarierea fondului de producție din grupa a II-a se urmărește o precizie de ± 10%, cîtă vreme pentru inventarierile în cadrul codrului grădinărit se impune o precizie mai ridicată, de 5 — 8%.

Pentru lucrările de cercetare științifică nu se pot da indicații privind precizia; aceasta depinde de scopul urmărit și de natura fenomenului studiat.

Valoarea u sau t se stabilește în funcție de probabilitatea cu care trebuie garantat rezultatul selecției. Pentru majoritatea cercetărilor din domeniul economiei forestiere se dovedește indicată probabilitatea de 95%. În unele situații, unde se urmărește un înalt grad de siguranță a rezultatelor, se consideră P = 99 sau chiar 99,9%.

Abaterea medie pătratică sau coeficientul de variație de cele mai multe ori sînt necunoscute. De aceea se procedează după cum urmează:

— se ia în calcul, în mod provizoriu, rezultatele unor cercetări anterioare efectuate în condiții similare;

— dacă asemenea cercetări lipsesc, se efectuează un eșantion restrîns, numai în scopul stabilirii cu aproximație a coeficientului de variație.

Pentru principalele caracteristici taxatorice ale arboretelor, ca primă orientare, se pot folosi coeficienții de variație prezentați la § 2.5. În cazul inventarierilor parțiale, o determinare aproximativă a volumului selecției se poate face pornind de la coeficienții de variație medii prezentați în tabelele 8.1 și 8.2 (G i u r g i u, 1964).

Tabelul 8.1

Coeficienți de variație medii în raport cu consistența, structura și clasa de omogenitate a arboretelor (Suprafața de sondaj de 500 m², arborete exploatabile)

Clasa de consistență	Structură echiană			Structura pluriană		
	Clasa de omogenitate					
	I	II	III	I	II	III
0,1—0,4	37	55	70	38	58	75
0,5—0,6	28	41	54	30	43	56
0,8 și peste	21	31	41	25	37	49

Scanned with OKEN Scanner

Tabelul 8.2

Coeficienţi de variaţie medii ai volumului pentru arborete echiene în cazul inventarierilor parţiale prin suprafeţe de probă circulare de 100 şi 200 m² (valori informative)

Consistenţa arboretelor	Mărimea suprafeţelor de probă, m²	
	100 (arborete tinere)	200 (arborete de vîrste mijlocii, neexploatabile)
0,1—0,4	71	—
0,5—0,7	59	48
0,8 şi peste	49	40

Pentru studiul proprietăţilor fizico-mecanice ale lemnului sînt utile următoarele valori orientative ale coeficienţilor de variaţie (L e o n t i e v, 1961) (în procente):

— greutatea specifică 10
— rezistenţa la compresiune paralel cu fibrele 13
— rezistenţa statică la încovoiere 15
— rezistenţa la tracţiune paralel cu fibrele 20
— rezistenţa la tracţiune perpendicular pe fibre 20

Pentru greutatea a 1000 seminţe se pot lua în considerare următoarele valori (prelucrare după C r i s t e s c u ş.a., 1957)

— răşinoase	10—15
— quercinee, tei, dîrmox	20—30
— păducel	15—20
— sînger, salbă moale, lemn cîinesc	20—25
— corn, glădiţă, salcîm	15
— frasin	20

În cazul caracteristicilor alternative, atunci cînd nu se cunoaşte valoarea dispersiei, volumul selecţiei se stabileşte iniţial pornind de la $p = 0{,}5$.

E x e m p l u l 4. Pentru determinarea greutăţii seminţelor de salcîm, cu o precizie de $\pm 2{,}5\%$ şi cu o probabilitate de 95%, urmează să efectuăm măsurători la

$$n = \frac{u^2 s_{\%}^2}{\Delta_{\%}^2} = \frac{1{,}96^2 \times 15^2}{2{,}5^2} = 144 \approx 150 \text{ probe}^*$$

Prelucrarea datelor obţinute în baza celor 150 probe analizate arată că eroarea standard a mediei aritmetice este de 0,23. Deci, pentru $P = 95\%$ stabilim următorul interval de încredere:

$22{,}80 \pm 1{,}96 \times 0{,}23 = 22{,}80 \pm 0{,}45$

Eroarea de 0,45, exprimată în procente (2%), este inferioară celei admise iniţial (2,5%). Rezultă deci că potrivit postcalculului efectuat volumul selecţiei a fost suficient. În caz contrar ar fi fost necesar să mărim volumul selecţiei prin măsurători suplimentare.

* S-a aplicat formula (8.11) deoarece colectivitatea generală este practic infinită. Din experimentări anterioare cunoaştem că $s\%$ este de cca. 15%.

Scanned with OKEN Scanner

8. 3. SELECȚIA ÎNTÎMPLĂTOARE REPETATĂ

Potrivit acestui procedeu fiecare unitate a colectivității generale poate fi extrasă și apoi din nou introdusă în colectivitatea generală, putînd participa la selecțiile următoare. Pentru determinarea erorii limită a mediei de selecție se folosesc formulele (8.3) și (8.4), iar pentru stabilirea aproximativă a volumului selecției rămîn valabile formulele (8.7) și (8.9).

Selecția repetată cere un volum mai mare de muncă. Este rar folosită în cercetările și lucrările de producție din sectorul economiei forestiere.

*

Atît selecția repetată cît și cea nerepetată se pot efectua și prin extragerea la întîmplare a unor ansambluri care pot fi considerate ca unități de observare.

În acest caz, al selecției pe serii, se pune problema extragerii la întîmplare a acestor unități-complexe formate din mai multe unități individuale. Fiecare serie extrasă este asimilată cu unitatea elementară, urmînd ca toate unitățile individuale ale fiecărei serii să fie înregistrate integral.

Pentru calculul erorii limită și al volumului selecției se pot folosi formulele cunoscute (8.1—8.11), cu mențiunea că, de data aceasta s^2 reprezintă dispersia dintre serii, n — numărul seriilor din colectivitatea de selecție, iar N — numărul seriilor din colectivitatea generală.

Pentru unele serii de volum mare este posibilă organizarea unei noi selecții întîmplătoare chiar în cadrul seriei.

8.4. SELECȚIA STRATIFICATĂ — NEREPETATĂ

In majoritatea cazurilor trebuie inventariate colectivități mai mult sau mai puțin eterogene. De pildă, fondul de producție al unui ocol silvic, care poate constitui obiectul unei inventarieri parțiale, variază sensibil de la arboret la arboret, în raport cu deosebirile staționale, de vîrstă, de îngrijire etc. Tot așa, procentul de împădurire pentru toate plantațiile efectuate în cadrul unui ocol silvic prezintă variații apreciabile din cauza deosebirilor de stațiune, proveniență și a altor factori. În cazul în care toate arboretele sînt considerate unități ale aceleiași colectivități generale, se impune folosirea unui procent ridicat de inventariere, fapt determinat de dispersia accentuată a fondului de producție pentru întreg fondul forestier luat în considerare.

In scopul reducerii volumului selecției, colectivitatea generală se desparte în mai multe straturi (grupe tipice), avînd în vedere principiile de bază ale grupării. În a doua etapă, considerînd fiecare strat ca o colectivitate independentă, se trece la efectuarea sondajului.

Astfel, arboretele considerate pot fi stratificate (grupate) pe mari unități staționale (tipologice), în care se grupează unitățile omogene din punct de vedere al structurii, productivității, compoziției specifice, consistenței. Această stratificare se poate realiza fie cu ajutorul fotogramelor, fie după o recunoaștere prealabilă a masivului de inventariat. În continuare, pentru

Scanned with OKEN Scanner

fiecare strat se aplică cel mai indicat procent de inventariere, avînd în vedere atît mărimea coeficienților de variație ai volumului, cît și considerentele de ordin economic.

Asemenea stratificări se impun și în alte lucrări de inventariere parțială din domeniul economiei forestiere (stabilirea procentului de reușită a culturilor forestiere; aprecierea stării fitosanitare a pădurilor; culegerea informațiilor privind fenomenul de uscare a anumitor specii forestiere, inventarierea resurselor micologice din fondul forestier, inventarierea efectivului de vînat etc.). În domeniul cercetării, problema stratificării prezintă, de asemenea, o deosebită importanță.

Eficacitatea stratificării se poate dovedi prin analiza dispersională (P r o d a n, 1961). Suma pătratelor corespunzătoare colectivității generale se compune din suma pătratelor în cadrul straturilor și din suma pătratelor dintre straturi. *Prin stratificare se ajunge la o omogenizare a colectivităților astfel formate; dispersia în cadrul fiecărui strat este relativ redusă, ceea ce, evident, asigură condiții mai bune pentru efectuarea unui sondaj avantajos.*

P r o d a n (1961), în baza datelor privind producția lemnului de fag din 215 ocoale silvice, prezintă următoarea situație (m^3):

Tabelul 8.3

Stratul	Suma volumelor	N_j	Media	Suma pătratelor	Dispersia s_j^2
1	14 300	140	102,1	1 223 957	8 805,45
2	25 190	50	503,8	893 178	18 228,12
3	28 100	25	1 124,0	1 944 200	81 008,33
Total	67 590	215	314,4	28 548 690	133 405,09

Colectivitatea generală, formată din 125 unități, a fost scindată în tre straturi avînd respectiv 140, 50 și 25 unități.

Analiza dispersională conduce la următoarele rezultate:

Împrăștierea	Suma pătratelor	Numărul gradelor de libertate	%
Intre straturi	24 487 346,6	2	85,8
In cadrul straturilor	4 061 335	212	14,2
Total	28 548 681,6	214	100

Rezultă că, din totalul sumei pătratelor, 86% revin dispersiei între cele trei straturi și numai 14% dispersiei în cadrul straturilor. Astfel, prin efectuarea stratificării, eroarea de reprezentativitate poate fi mult redusă ca urmare a restrîngerii dispersiei în cadrul straturilor.

Scanned with OKEN Scanner

Mărimea erorilor de selecție depinde în mare măsură de modul în care se repartizează pe straturi unitățile colectivității de selecție. Această repartizare poate fi: a) simplă, b) proporțională, c) optimă.

a) În cazul *sondajului stratificat simplu*, eroarea mediei de selecție rezultă din formula:

$$s_{\bar{x}} = \sqrt{\Sigma \frac{N_j^2}{N^2}\left(1 - \frac{n_j}{N_j}\right)\frac{s_j^2}{n_j}} \qquad (8.12)$$

unde:

N — volumul colectivității generale;
N_j — numărul unităților din fiecare strat;
n_j — numărul unităților extrase din fiecare strat;
s_j — abaterea medie pătratică a fiecărui strat.

Din fiecare strat se extrage aproximativ același număr de unități. În exemplul din tabelul 8.4 s-au extras $6 + 6 + 5 = 17$ unități.

Tabelul 8.4

Folosirea sondajului stratificat simplu la inventarierea producției lemnului de fag (Prodan, *1961*)

Stratul	N_j	n_j	media	$\frac{N_j}{N}$	$\left(\frac{N_j}{N}\right)^2$	$\frac{n_j}{N_j}$	$1-\frac{n_j}{N_j}$	s_j^2	$\frac{s_j^2}{n_j}$	$\left(\frac{N_j}{N}\right)^2\left(1-\frac{n_j}{N_j}\right)\frac{s_j^2}{n_j}$
1	140	6	140	0,65	0,423	0,043	0,957	14 200	2 366	957
2	50	6	520	0,23	0,053	0,120	0,880	60 200	10 033	467
3	25	5	1 250	0,12	0,014	0,100	0,900	85 120	17 024	204
Σ	215	17	387	1,00	—	—	—	—	—	1 628

Introducînd în formula (8.12) suma obținută în tabelul 8.4, rezultă:

$$s_{\bar{x}} = \sqrt{1\,628} = 40{,}4 \text{ m}^3$$

Așa încît, eroarea limită pentru o probabilitate de 95% va fi: $2{,}12 \times 40{,}4 = 85{,}6$ m³.

Deci, volumul mediu va fi cuprins în intervalul 387 ± 86 ceea ce înseamnă că volumul mediu pentru întreaga colectivitate de 215 unități nu va fi mai mic de 301 și nici mai mare de 473 m³ și aceasta se poate afirma cu o siguranță statistică de 95%. Pentru colectivitatea generală limitele vor fi: $387 \times 215 \pm 86 \times 215$ sau $83\,205 \pm 1\,849$.

b) În cazul *sondajului stratificat proporțional*, din fiecare strat se extrag atîtea unități astfel încît raportul dintre numărul lor și volumul stratului din care s-au extras să fie egal cu raportul dintre volumul total al selecției și cel al colectivității generale. De data aceasta volumul selecției exprimat în procente reprezintă o mărime practic constantă pentru toate straturile.

Scanned with OKEN Scanner

Eroarea limită este dată de formula:

$$s_{\bar{x}} = \sqrt{\frac{1-\frac{n}{N}}{n} \sum_{j=1}^{K} \frac{N_j}{N} s_j^2} \qquad (8.13)$$

unde:

n/N — raportul dintre volumul selecției și volumul colectivității generale, păstrat același pentru toate straturile formate;

K — numărul straturilor formate.

Pentru același exemplu, adoptînd $n/N = 0{,}08$ (sau $n = 17$) extragem 11 unități din primul strat, 4 unități din al doilea și 2 unități din ultimul strat. Rezultatele obținute sînt centralizate în tabelul 8.5.

Tabelul 8.5

Stratul	N_j	n_j	$\frac{N_j}{N}$	s_j^2	$\frac{N_j}{N} s_j^2$	Media
1	140	11	0,65	10 825	7 036	96
2	50	4	0,23	50 300	11 569	544
3	25	2	0,12	72 200	8 664	1 480
Σ	215	17	1,00	—	27 269	379

Aplicînd formula (8.13) obținem:

$$s_{\bar{x}} = \sqrt{\frac{1-\frac{17}{215}}{17} \times 27\,269} = \sqrt{1\,478} = 38{,}4 \text{ m}^3$$

În cadrul unei siguranțe probabilistice de 95%, limitele intervalului de încredere vor fi:

— pentru medie

$$379 \pm 2{,}12 \times 38 = 298 \text{ și } 460 \text{ m}^3$$

— pentru colectivitatea generală

$$81\,560 \pm 2{,}12 \times 8\,213 = 64\,149 \text{ și } 98\,971 \text{ m}^3$$

Se constată că sondajul stratificat proporțional, la același volum al colectivității de selecție și la aceeași siguranță statistică, este mai eficient decît sondajul stratificat simplu, eroarea de reprezentabilitate fiind mai mică. Explicația constă în faptul că în acest caz înăuntrul fiecărui strat format se obține o dispersie mai redusă.

Scanned with OKEN Scanner

c) *Sondajul stratificat optim* asigură o eficiență maximă a sondajului. Volumul selecției pe straturi (n_j) este stabilit potrivit expresiei:

$$n_j = \frac{n\, N_j\, s_j}{\sum_{j=1}^{K} N_j s_j} \tag{8.14}$$

Abaterea medie pătratică a mediei de selecție va fi:

$$s_{\bar{x}} = \sqrt{\frac{1}{N^2}\left[\frac{\left(\sum_{j=1}^{K} N_j\, s_j\right)^2}{n} - \sum_{j=1}^{K} N_j\, s_j^2\right]} \tag{8.15}$$

În acest caz se impune cunoașterea abaterilor medii pătratice pe straturi. Pentru exemplul considerat, avem:

$$n_1 = \frac{17\times140\times94}{140\times94+50\times135+25\times285} = \frac{22\,3720}{27\,035} = 8$$

$$n_2 = \frac{17\times50\times135}{27\,035} = 4$$

$$n_3 = \frac{17\times25\times285}{27\,035} = 5$$

Rezultatele selecției efectuate prin extragerea din fiecare strat a numărului de unități stabilit sînt prezentate în tabelul 8.6.

Tabelul 8.6

Stratul	N_j	Media după datele selecției	s_j^2 după datele selecției	s_j după datele selecției	$N_j\, s_j^2$	$N_j\, s_i$
1	140	115	11 657	108	1 631 980	15 120
2	50	545	50 300	224	2 515 000	11 200
3	25	1 140	118 900	345	2 972 500	8 625
Σ	215	334	—	—	7 119 480	34 945

Datele din acest tabel le introducem în formula (8.15) și obținem:

$$s_{\bar{x}} = \sqrt{\frac{1}{215^2}\left[\frac{34\,945^2}{17} - 7\,119\,480\right]} = 37{,}4 \text{ m}^3$$

La alte sondaje stratificate, efectuate în condiții identice, media de selecție (care la această probă a dat 334 m³) poate lua alte valori cuprinse în următorul interval ($P = 95\%$, $t = 2{,}12$, anexa 4): $334 \pm 2{,}12\times37 = 255$ și 423.

Scanned with OKEN Scanner

Pentru colectivitatea generală volumul va fi cuprins între 54 803 și 88 897 m³.

Folosind cele trei procedee de repartizare a unităților selecției, în cadrul aceluiași volum al colectivității de selecție și la același nivel de semnificație, au rezultat următoarele erori de reprezentativitate:

— sondajul statificat simplu	40,4 m³
— sondajul stratificat proporțional	38,4 m³
— sondajul stratificat optim	37,4 m³

După cum se vede, sondajul stratificat optim conduce la cele mai bune rezultate sub raportul preciziei.

Folosind procedeul selecției stratificate, va trebui permanent avut în vedere necesitatea formării unor straturi care după conținutul lor să fie cît mai omogene din punct de vedere al caracteristicilor luate în considerare. Procedînd astfel, înlăturăm influența *variației dintre straturi* asupra erorii limită.

În scopul determinării volumului selecției, se pot folosi formulele:

— pentru caracteristici nealternative

$$n = \frac{u^2 \overline{s}^2}{\Delta_x^2 + \frac{u^2 \overline{s}^2}{N}} \tag{8.16}$$

unde:

$\overline{s}^2$ — media ponderată a dispersiilor celor K straturi formate;

— pentru caracteristici alternative

$$n = \frac{u^2 p (1 - p)}{\Delta_p^2 + u^2 p (1 - p)} \tag{8.17}$$

În continuare, unitățile se repartizează pe straturi potrivit procedeului adoptat.

Acest procedeu, al selecției stratificate, este frecvent folosit în S.U.A. și Canada, pentru inventarierea fondului forestier (H a s e l, A. A., 1954). Stratificarea se realizează prin procedee moderne aerofotogrametrice. Pentru fiecare strat format se organizează apoi cercetări de teren pe suprafețe de probă, urmărind în principal evaluarea fondului de producție din punctul de vedere al mărimii creșterii și calității lui. Mai recent, acest procedeu a luat o amploare deosebită, mai ales pentru inventarierea fondului forestier pe suprafețe intense (L o e t s c h-H a l l e r, 1964).

8. 5. SELECȚIA MECANICĂ

Sondajul sistematic, cum mai este denumită selecția mecanică, constă în alegerea unităților colectivității de selecție, după o schemă fixă. De obicei se procedează astfel: volumul colectivității generale se împarte la volumul colectivității de selecție; rezultatul obținut, numit *pas de numărare*,

arată ordinea în care trebuie extrase unităţile. În cazul cînd volumul selecţiei reprezintă 10% din volumul colectivităţii generale, pasul de numărare va fi 100 : 10 = 10; prin urmare, în colectivitatea de selecţie se vor include, spre exemplu, prima valoare, a 10-a, a 20-a, a 30-a ş.a.

Acest procedeu s-a dovedit corespunzător în special pentru inventarierea fondului de producţie în cadrul lucrărilor de amenajare a pădurilor, precum şi în alte lucrări de producţie. În lucrările de inventariere pe suprafeţe mari, selecţia mecanică s-a dovedit a fi cu mult mai indicată în comparaţie cu selecţia întîmplătoare. De regulă, selecţia mecanică este precedată de stratificarea corespunzătoare a colectivităţii generale.

Selecţia mecanică prezintă următoarele avantaje:

— prinde mai bine toate părţile colectivităţii generale;

— asigură o mai uşoară organizare a lucrărilor de inventariere;

— de cele mai multe ori selecţia sistematică, datorită simplităţii în efectuarea ei, prezintă o mai mare siguranţă, fiind în acelaşi timp mai ieftină.

În unele situaţii selecţia sistematică este mai favorabilă chiar decît selecţia stratificată. Prezintă însă două dezavantaje:

— o oarecare nesiguranţă din cauza unor tendinţe necunoscute, mai ales atunci cînd elementele colectivităţii generale sînt afectate de o periodicitate oarecare;

— unele dificultăţi de ordin teoretic privind calculul corect al erorilor de reprezentativitate (S n e d e c o r, 1957).

În cazul selecţiei sistematice, pentru determinarea erorilor probabilistice se aplică formulele de calcul folosite pentru selecţia simplă repetată

Folosind formulele stabilite pentru selecţia simplă, mărim precizia selecţiei, deoarece erorile date de selecţia mecanică sînt mai mici decît cele ale sondajului simplu repetat, dar mai mari decît ale sondajului simplu nerepetat (M i h o c şi U r s e a n u, 1962).

Pentru lucrările de inventariere parţială a fondului forestier prin procedee sistematice prezintă un deosebit interes practic problema mărimii suprafeţelor de probă. În vederea elucidării acestei probleme se poate folosi teoria scindării dispersiei (P r o d a n, 1961).

Un exemplu de selecţie sistematică îl constituie procedeul inventarierii fondului de producţie prin benzi de probă (P o p e s c u-Z e l e t i n, 1947).

8. 6. SONDAJE IN DOUĂ SAU MAI MULTE FAZE

Sondajul în două faze se realizează alegînd în prima fază h straturi dintr-o colectivitate generală împărţită în K straturi ($h < K$). În a doua fază se efectuează o nouă selecţie în cadrul fiecărui strat inclus în selecţie (fig. 8.1).

Acest procedeu se foloseşte atunci cînd se impune obţinerea unor informaţii în timp scurt cu cheltuieli reduse.

Scanned with OKEN Scanner

Selecția în două faze poate fi folosită și în alte scopuri. De pildă, în prima fază se înregistrează toate unitățile colectivității de selecție, iar în faza a doua se cercetează numai o parte din colectivitatea de selecție, urmărind înregistrarea unor caracteristici suplimentare. Spre exemplu, în prima fază se pot înregistra diametrele arborilor din anumite suprafețe de probă amplasate sistematic, iar în a doua fază urmează măsurarea înălțimilor și a creșterilor în diametru numai la o parte din arborii fiecărei suprafețe de probă amplasate.

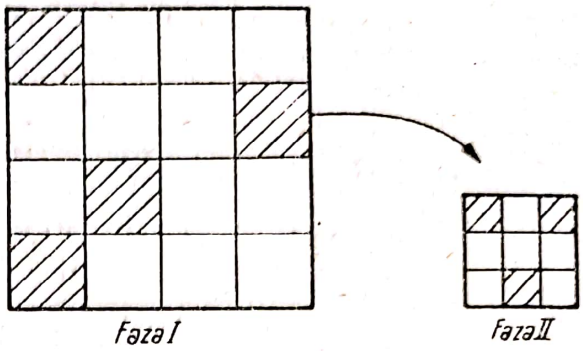


Fig. 8.1. Schema selecției în două faze

În cadrul sondajului în trei faze în prima etapă, formăm, spre exemplu, o colectivitate de selecție prin alegerea la întîmplare a unor arborete, respectînd însă principiul stratificării; în a doua fază se formează *subcolectivități de selecție*, alegînd arbori de probă; în a treia fază — în cadrul subcolectivităților constituite — se trece la alegerea, de pildă, a frunzelor, ramurilor sau a conurilor ce urmează a fi în detaliu examinate.

În figura 8.2 se poate urmări schema selecției în trei faze.

În comparație cu celelalte procedee de selecție, sondajul în două sau mai multe faze este mai puțin precis, fapt explicabil dacă avem în vedere că, pe măsură ce se trece la formarea unor subcolectivități de selecție de ordin inferior, sporește eroarea de reprezentativitate. De aceea numărul exagerat al fazelor trebuie evitat.

Se demonstrează că eroarea mediei de selecție a sondajului efectuat în două faze este dată de formula:

$$s_{\bar{x}} = \sqrt{\frac{s_1^2}{n_1}\left(1-\frac{n_1}{N_1}\right)+\frac{s_2^2}{n_1 n_2}\left(1-\frac{n_2}{N_2}\right)} \qquad (8.18)$$

unde:

s_1^2 și s_2^2 — caracterizează împrăștierea în interiorul unităților de primul tip și respectiv împrăștierea în interiorul unităților de tipul doi;

n și n_2 — numărul unităților de tipul I și respectiv numărul unităților de tipul II incluse în selecție;

N_1 — numărul unităților de tipul II care compun colectivitatea studiată;

N_2 — numărul de unități de al doilea tip conținute de o unitate de tipul I.

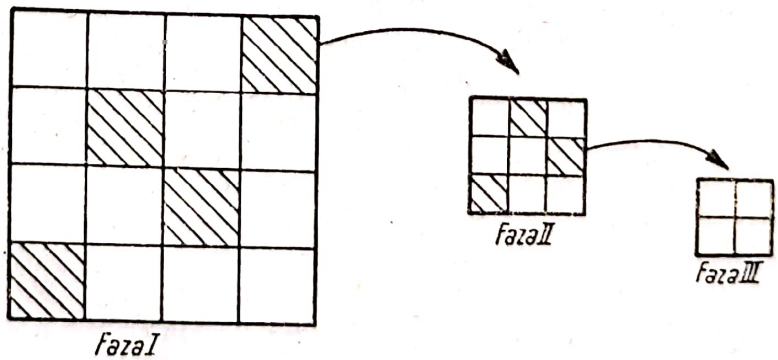


Fig. 8.2. Schema selecției în trei faze

Scanned with OKEN Scanner

Pentru organizarea lucrărilor prezintă interes raportul dintre volumele de selecţie ale celor două faze, luînd în considerare aspecte de ordin economic. Se demonstrează că

$$Q = c_1 n_1 + c_2 n_1 n_2 \tag{8.19}$$

unde:

Q — cheltuielile totale ale selecţiei;
c_1 — cheltuielile primei faze, raportate la unitatea de selecţie de ordinul întîi;
c_2 — cheltuielile celei de a doua faze, raportate la unitatea de selecţie de ordinul doi;
n_1 — numărul unităţilor din prima fază a selecţiei;
n_2 — numărul unităţilor din a doua fază a selecţiei.

Pentru determinarea lui n_2 se recomandă formula:

$$n_2 = \sqrt{\frac{c_1 s_2^2}{c_2 s_1^2}}\left(1 - \frac{s_2^2}{N_2 s_1^2}\right)^{-\frac{1}{2}} \tag{8.20}$$

În cazul cînd colectivitatea generală este infinită sau atunci cînd N este foarte mare, numărul unităţilor de tipul doi se obţine din relaţia

$$n_2 = \frac{s_2}{s_1}\sqrt{\frac{c_1}{c_2}} \tag{8.21}$$

De observat că n_2 nu depinde de costul cercetării, ci numai de raportul cheltuielilor c_1 şi c_2 şi al dispersiilor s_1^2 şi s_2^2. În continuare se trece la determinarea lui n_1, folosind formula (8.20). În acest caz trebuie să se stabilească, cel puţin cu aproximaţie, costul total al selecţiei (Q). Procedee asemănătoare se prezintă şi pentru selecţia în trei faze (S n e d e c o r, 1957).

Sondajul în două sau mai multe faze prezintă o deosebită importanţă pentru planificarea şi organizarea multor lucrări privind: inventarierea fondului de producţie şi a creşterii acestuia, inventarierea potenţialului staţional şi al condiţiilor ecologice pe suprafeţe mari forestiere, problema daunelor aduse fondului forestier de atacurile de dăunători etc. Prezintă de asemenea interes pentru multe cercetări de laborator.

*

Deosebit de eficientă este aplicarea metodei selective în domeniul dirijării proceselor bio-producţiei forestiere, la controlul operativ al proceselor de fabricaţie în întreprinderile forestiere, la verificarea calităţii loturilor de produse forestiere (seminţe, sortimente de lemn brut, cherestea etc.) la cercetarea utilizării timpului de lucru al muncitorilor, la îndeplinirea normelor de lucru etc. În acest scop sînt utile lucrările elaborate de R a n c u şi T ö v i s s i (1963, 1964), precum şi articolele publicate în colecţia „Calcul economic" (M a n e a M ă n e s c u şi col., 1964).

Indicaţii suplimentare privind metoda selectivă se pot obţine şi din lucrările semnate de M i h o c şi U r s e a n u (1962), B i j i şi S t o i c h i ţ ă (1957), C o c h r a n (1953); Y a t e s (1954), S n e d e c o r (1957), H a l d (1952) ş.a.

Teoria și aplicațiile practice ale metodei selective la inventarierea fondului forestier sînt pe larg prezentate de L o e t s c h și H a l l e r (1964), P r o d a n (1961), S c h u m a c h e r și C h a p m a n (1954), J e f f e r s (1960), H u s c h (1963).

Problematica inventarierii efectivului de animale sălbatice prin metodele statisticii matematice este descrisă de S c a t t e r g o o d (1954). Pentru probleme de genetică este utilă lucrarea lui K e m p t h o r n e (1963).

Scanned with OKEN Scanner

Anexe

Scanned with OKEN Scanner

Valorile $x^4 n_j$

n_j	$x=2$	$x=3$	$x=4$	$x=5$	$x=6$	$x=7$
1	16	81	256	625	1296	2401
2	32	162	512	1250	2592	4802
3	48	243	768	1875	3888	7203
4	64	324	1024	2500	5184	9604
5	80	405	1280	3125	6480	12005
6	96	486	1536	3750	7776	14406
7	112	567	1792	4375	9072	16807
8	128	648	2048	5000	10368	19208
9	144	729	2304	5625	11664	21609
10	160	810	2560	6250	12960	24010
11	176	891	2816	6875	14256	26411
12	192	972	3072	7500	15552	28812
13	208	1053	3328	8125	16848	31213
14	224	1134	3584	8750	18144	33614
15	240	1215	3840	9375	19440	36015
16	256	1296	4096	10000	20736	38416
17	272	1377	4352	10625	22032	40817
18	288	1458	4608	11250	23328	43218
19	304	1539	4864	11875	24624	45619
20	320	1620	5120	12500	25920	48020
21	336	1701	5376	13125	27216	50421
22	352	1782	5632	13750	28512	52822
23	368	1863	5888	14375	29808	55223
24	384	1944	6144	15000	31104	57624
25	400	2025	6400	15625	32400	60025
26	416	2106	6656	16250	33696	62426
27	432	2187	6912	16875	34992	64827
28	448	2268	7168	17500	36288	67228
29	464	2349	7424	18125	37584	69629
30	480	2430	7680	18750	38880	72030
31	496	2511	7936	19375	40176	74431
32	512	2592	8192	20000	41472	76832
33	528	2673	8448	20625	42768	79233
34	544	2754	8704	21250	44064	81634
35	560	2835	8960	21875	45360	84035
36	576	2916	9216	22500	46656	86436
37	592	2997	9472	23125	47952	88837
38	608	3078	9728	23750	49248	91238
39	624	3159	9984	24375	50544	93639
40	640	3240	10240	25000	51840	96040
41	656	3321	10496	25625	53136	98441
42	672	3402	10752	26250	54432	100842
43	688	3483	11008	26875	55728	103243
44	704	3564	11264	27500	57024	105644
45	720	3645	11520	28155	58320	108045
46	736	3726	11776	28750	59616	110446
47	752	3807	12032	29375	60912	112847
48	768	3888	12288	30000	62208	115248
49	784	3969	12544	30625	63504	117649
50	800	4050	12800	31250	64800	120050

Pentru frecvenţe mai mari de 50, se procedează la scindarea lui n_j, însumînd rezultate parţiale: $n = 55, = 50 + 5$; $x = 2$; $x^4 n_j = 800 + 80 = 880$.

Scanned with OKEN Scanner

Valorile funcției de frecvență ale distribuției normale $f(u)$ și ale derivatelor $f^{III}(u)$ și $f^{IV}(u)$*

u	$f(u)$	$f^{III}(u)$	$f^{IV}(u)$	u	$f(u)$	$f^{III}(u)$	$f^{IV}(u)$
0,00	0,39894	+0,00000	+1,19683	0,50	0,35207	+0,48409	+0,55010
02	39886	02393	19563	52	34849	49465	50556
04	39862	04781	19204	54	34482	50431	46048
06	39822	07159	18608	56	34105	51306	41497
08	39767	09524	17775	58	33718	52091	36913
10	39695	11869	16708	60	33322	52783	32309
12	39608	14190	15410	62	32918	53383	27696
14	39505	16484	13885	64	32506	53891	23085
16	39387	18744	12137	66	32086	54306	18486
18	39253	20968	10170	68	31659	54630	13912
20	39104	23150	07990	70	31225	54863	09371
22	38940	25286	05004	72	30785	55005	04874
24	38762	27373	03018	74	30339	55058	+0,00433
26	38568	29405	1,00238	76	29887	55023	03944
28	38361	31381	97273	78	29430	54901	08248
30	38139	33295	94130	80	28969	54694	12468
32	37903	35145	90819	82	28504	54403	16597
34	37654	36927	87348	84	28034	54031	20626
36	37391	38638	83726	86	27562	53579	24546
38	37115	40275	79963	88	27086	53049	28351
40	36827	41835	76070	90	26609	52445	32034
42	36526	43317	72056	92	26129	51769	35587
44	36213	44717	67932	94	25647	51023	39005
46	35889	46034	63709	96	25164	50210	42283
48	35553	47265	59398	98	24681	49332	45414
0,50	0,35207	+0,48409	+0,55010	1,00	0,24197	+0,48394	—0,48394
1,00	0,24197	+0,48394	—0,48394	1,50	0,12952	+0,14571	—0,70425
02	23713	47398	51220	52	12566	13172	69423
04	23230	46346	53887	54	12188	11795	68314
06	22747	45243	56393	56	11816	10440	67104
08	22265	44092	58734	58	11450	09111	65799
10	21785	42895	60909	60	11092	07809	64405
12	21307	41657	62917	62	10741	06535	62928
14	20831	40380	64755	64	10396	05292	61375
16	20357	39067	66425	66	10059	04081	59751
18	19886	37724	67924	68	09728	02903	58063
20	19419	36352	69255	70	09405	01759	56316
22	18954	34955	70417	72	09089	00650	54516
24	18494	33536	71411	74	08780	—0,00422	52671
26	18037	32099	72241	76	08478	01456	50785
28	17585	30648	72907	78	08183	02453	48865
30	17137	29184	73413	80	07895	03411	46915
32	16694	27712	73760	82	07614	04329	44943
34	16256	26235	73953	84	07341	05208	42953
36	15822	24755	73995	86	07074	06047	40950
38	15395	23276	73890	88	06814	06846	38940
40	14973	21800	73642	90	06562	07605	36928

* După MITROPOLSKI (1961)

Scanned with OKEN Scanner

(continuare)

ANEXA 2

u	$f(u)$	$f^{III}(u)$	$f^{IV}(u)$	u	$f(u)$	$f^{III}(u)$	$f^{IV}(u)$
42	14556	20331	73256	92	06316	08323	34918
44	14146	18871	72736	94	06077	09002	32916
46	13742	17423	72087	96	05844	09640	30925
48	13344	15988	71315	98	05618	10239	28950
1,50	0,12952	+0,14571	—0,70425	2,00	0,05399	—0,10798	—0,26996
2,00	0,05399	—0,10798	—0,26996	2,50	0,01753	—0,14242	+0,07997
02	05186	11319	25064	52	01667	14075	08710
04	04980	11801	23160	54	01585	13894	09372
06	04780	12245	21287	56	01506	13700	09982
08	04586	12653	19448	58	01431	13495	10542
10	04398	13024	17646	60	01358	13279	11053
12	04217	13359	15883	62	01289	13053	11517
14	04041	13659	14162	64	01223	12818	11935
16	03871	13926	12486	66	01160	12576	12308
18	03706	14159	10856	68	01100	12326	12638
20	03547	14360	09274	70	01042	12071	12926
22	03394	14530	07743	72	00987	11810	13174
24	03246	14670	06263	74	00935	11544	13383
26	03103	14781	04835	76	00885	11274	13555
28	02965	14864	03461	78	00837	11002	13691
30	02833	14920	02141	80	00792	10727	13793
32	02705	14950	00877	82	00748	10450	13863
34	02582	14955	+0,00332	84	00707	10173	13902
36	02463	14937	01485	86	00668	09895	13912
38	02349	14896	02582	88	00631	09616	13894
40	02239	14834	03623	90	00595	09339	13850
42	02134	14752	04608	92	00562	09063	13782
44	02033	14650	05537	94	00530	08788	13691
46	01936	14531	06411	96	00499	08515	13579
48	01842	14394	07231	98	00470	08245	13446
2,50	0,01753	—0,14242	+0,07997	3,00	0,00443	—0,07977	+0,13296
3,00	0,00443	—0,07977	+0,13296	3,50	0,00087	—0,02825	+0,06943
05	00381	07323	12847	55	00073	02494	06308
10	00327	06694	12313	60	00061	02194	05703
15	00279	06093	11714	65	00051	01923	05133
20	00238	05523	11066	70	00042	01680	04599
25	00203	04987	10387	75	00035	01463	04103
30	00172	04485	09690	80	00029	01269	03646
35	00146	04018	18987	85	00024	01098	03226
40	00123	03586	08290	90	00020	00946	02842
45	00104	03189	07606	95	00016	00813	02495
3,50	0,00087	—0,02825	+0,06943	4,00	0,00013	—0,00696	+0,02181

Scanned with OKEN Scanner

ANEXA 3

VALORILE PROBABILITĂȚILOR $q = 1 - 2\Phi(u)$ și $P = 2\Phi(u) = 1 - q$, corespunzătoare distribuției normale

u în funcție de q			q în funcție de u		
q	P	u	u	q	P
în procente				în procente	
100	0,00	0,0000	0,0	100,000	0,000
95	5	0,0627	0,2	84,148	15,852
90	10	0,1257	0,4	68,916	31,084
85	15	0,1891	0,6	54,851	45,149
80	20	0,2533	0,8	42,371	57,629
75	25	0,3186	1,0	31,731	68,269
70	30	0,3853	1,2	23,014	76,986
65	35	0,4538	1,4	16,151	83,849
60	40	0,5244	1,6	10,960	89,040
55	45	0,5978	1,8	7,186	92,814
50	50	0,6795	2,0	4,555	95,445
45	55	0,7554	2,2	2,781	97,219
40	60	0,8416	2,4	1,640	98,360
35	65	0,9346	2,6	0,932	99,068
30	70	1,0364	2,8	0,511	99,489
25	75	1,1503	3,0	0,270	99,730
20	80	1,2816	3,2	0,137	99,863
15	85	1,4395	3,4	0,067	99,933
10	90	1,6449	3,6	0,032	99,968
5	95	1,9600	3,8	0,014	99,986
1	99	2,5758	4,0	0,006	99,994
0,27	99,73	3,0000			
0,1	99,9	3,2905			
0,01	99,99	3,8906			

Scanned with OKEN Scanner

ANEXA 4

VALORILE DISTRIBUȚIEI STUDENT

f	Nivelul de semnificație, %				
	10	5	2	1	0,1
1	6,31	12,71	31,82	63,66	636,2
2	2,92	4,30	6,97	9,93	31,60
3	2,35	3,18	4,54	5,84	12,94
4	2,13	2,78	3,75	4,60	8,61
5	2,02	2,57	3,37	4,03	6,86
6	1,94	2,45	3,14	3,70	5,96
7	1,90	2,37	3,00	3,50	5,40
8	1,86	2,30	2,90	3,36	5,04
9	1,83	2,26	2,82	3,25	4,78
10	1,81	2,23	2,76	3,17	4,59
11	1,80	2,20	2,72	3,11	4,49
12	1,78	2,18	2,68	3,06	4,32
13	1,77	2,18	2,65	3,06	4,12
14	1,76	2,14	2,62	2,98	4,14
15	1,75	2,13	2,60	2,95	4,07
16	1,75	2,12	2,58	2,92	4,02
17	1,74	2,11	2,57	2,90	3,97
18	1,73	2,10	2,55	2,88	3,92
19	1,73	2,09	2,54	2,86	3,88
20	1,72	2,09	2,53	2,85	3,85
21	1,72	2,08	2,52	2,83	3,82
22	1,72	2,07	2,51	2,82	3,79
23	1,71	2,07	2,50	2,81	3,77
24	1,71	2,06	2,49	2,80	3,75
25	1,71	2,06	2,49	2,79	3,72
26	1,71	2,06	2,48	2,78	3,71
27	1,70	2,05	2,47	2,77	3,69
28	1,70	2,05	2,47	2,76	3,67
29	1,70	2,05	2,46	2,76	3,66
30	1,70	2,04	2,46	2,75	3,65
40	1,68	2,02	2,42	2,70	3,55
60	1,67	2,00	2,39	2,66	3,36
120	1,66	1,98	2,36	2,62	3,37
∞	1,65	1,96	2,33	2,58	3,29

Scanned with OKEN Scanner

DISTRIBUȚIA F

f_2 \ f_1	Nivelul de semnificație 0,05 sau 5%								
	1	2	3	4	5	6	12	24	∞
1	164,4	199,5	215,7	224,6	230,2	234,0	244,9	249,0	254,3
2	18,5	19,2	19,2	19,3	19,3	19,3	19,4	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7	5,6	4,7	4,4	4,1	4,0	3,9	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
11	4,8	4,0	3,6	3,4	3,2	3,1	2,8	2,6	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,7	2,5	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,6	2,4	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,5	2,3	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,4	2,2	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,4	2,2	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,3	2,1	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,9
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
22	4,3	3,4	3,1	2,8	2,7	2,6	2,2	2,0	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,2	2,0	1,7
26	4,2	3,4	3,0	2,7	2,6	2,5	2,2	2,0	1,7
28	4,2	3,3	3,0	2,7	2,6	2,4	2,1	1,9	1,7
30	4,2	3,3	2,9	2,7	2,5	2,4	2,1	1,9	1,6
40	4,1	3,2	2,9	2,6	2,5	2,3	2,0	1,8	1,5
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
120	3,9	3,1	2,7	2,5	2,3	2,2	1,8	1,6	1,3
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

f_1 — numărul gradelor de libertate al dispersiei celei mai mari;
f_2 — numărul gradelor de libertate al dispersiei celei mai mici.

Scanned with OKEN Scanner

(continuare)

ANEXA 5

f_2 \ f_1	Nivelul de semnificație 0,025 sau 2,5%*								
	1	2	3	4	5	6	12	14	∞
1	648	800	864	900	922	937	977	997	1018
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,3	14,1	13,9
4	12,2	10,6	10,0	9,6	9,4	9,2	8,8	8,5	8,3
5	10,0	8,4	7,8	7,4	7,2	7,0	6,5	6,3	6,0
6	8,8	7,3	6,6	6,2	6,0	5,8	5,4	5,1	4,8
7	8,1	6,5	5,9	5,5	5,3	5,1	4,7	4,4	4,1
8	7,6	6,1	5,4	5,1	4,8	4,6	4,2	4,0	3,7
9	7,2	5,7	5,1	4,7	4,5	4,3	3,9	3,6	3,3
10	6,9	5,5	4,8	4,5	4,2	4,1	3,6	3,4	3,1
11	6,7	5,3	4,6	4,3	4,0	3,9	3,4	3,2	2,9
12	6,6	5,1	4,5	4,1	3,9	3,7	3,3	3,0	2,7
13	6,4	5,0	4,4	4,0	3,8	3,6	3,5	2,9	2,6
14	6,3	4,9	4,2	3,9	3,7	3,5	3,0	2,8	2,5
15	6,2	4,8	4,2	3,8	3,6	3,4	3,0	2,7	2,4
16	6,1	4,7	4,1	3,7	3,5	3,3	2,9	2,6	2,3
17	6,0	4,6	4,0	3,7	3,4	3,3	2,8	2,6	2,2
18	6,0	4,6	4,0	3,6	3,4	3,2	2,8	2,5	2,2
19	5,9	4,5	3,9	3,6	3,3	3,2	2,7	2,5	2,1
20	5,9	4,5	3,9	3,5	3,3	3,1	2,7	2,4	2,1
22	5,8	4,4	3,8	3,4	3,2	3,0	2,6	2,3	2,0
24	5,7	4,3	3,7	3,4	3,2	3,0	2,5	2,3	1,9
26	5,7	4,3	3,7	3,3	3,1	2,9	2,5	2,2	1,9
28	5,6	4,2	3,6	3,3	3,1	2,9	2,5	2,2	1,8
30	5,6	4,2	3,6	3,3	3,0	2,9	2,4	2,1	1,8
40	5,4	4,1	3,5	3,1	2,9	2,7	2,3	2,0	1,6
60	5,3	3,9	3,3	3,0	2,8	2,6	2,2	1,9	1,5
120	5,1	3,8	3,2	2,9	2,7	2,5	2,1	1,8	1,3
	5,0	3,7	3,1	2,8	2,6	2,4	1,9	1,6	1,0

* Pentru compararea a două dispersii (§ 4.3) această tabelă corespunde nivelului de semnificație de 0,05 sau 5%.

Scanned with OKEN Scanner

(continuare)

ANEXA 5

f_2 \ f_1	Nivelul de semnificație 0,01 sau 1%									
	1	2	3	4	5	6	8	12	24	∞
1	4052	4999	5403	5625	5764	5859	5981	6106	6234	6366
2	98,5	99,0	99,2	99,3	99,3	99,4	99,3	99,4	99,5	99,5
3	34,1	30,8	29,5	28,7	28,2	27,9	27,5	27,1	26,6	26,1
4	21,2	18,0	16,7	16,0	15,5	15,2	14,8	14,4	13,9	13,5
5	16,3	13,3	12,1	11,4	11,0	10,7	10,3	9,9	9,5	9,0
6	13,7	10,9	9,8	9,2	8,8	8,5	8,1	7,7	7,3	6,9
7	12,3	9,6	8,5	7,9	7,5	7,2	6,8	6,5	6,1	5,7
8	11,3	8,7	7,6	7,0	6,6	6,4	6,0	5,7	5,3	4,9
9	10,6	8,0	7,0	6,4	6,1	5,8	5,5	5,1	4,7	4,3
10	10,0	7,6	6,6	6,0	5,6	5,4	5,1	4,7	4,3	3,9
11	9,7	7,2	6,2	5,7	5,3	5,1	4,7	4,4	4,0	3,6
12	9,3	6,9	6,0	5,4	5,1	4,8	4,5	4,2	3,8	3,4
13	9,1	6,7	5,7	5,2	4,9	4,6	4,3	4,0	3,6	3,2
14	8,9	6,5	5,6	5,0	4,7	4,5	4,1	3,8	3,4	3,0
15	8,7	6,4	5,4	4,9	4,6	4,3	4,0	3,7	3,3	2,9
16	8,5	6,2	5,3	4,8	4,4	4,2	3,9	3,6	3,2	2,8
17	8,4	6,1	5,2	4,7	4,3	4,1	3,8	3,5	3,1	2,7
18	8,3	6,0	5,1	4,6	4,3	4,0	3,7	3,4	3,0	2,6
19	8,2	5,9	5,0	4,5	4,2	3,9	3,6	3,3	2,9	2,5
20	8,1	5,9	4,9	4,4	4,1	3,9	3,6	3,2	2,9	2,4
22	7,9	5,7	4,8	4,3	4,0	3,8	3,5	3,1	2,8	2,3
24	7,8	5,6	4,7	4,2	3,9	3,7	3,3	3,0	2,7	2,2
26	7,7	5,5	4,6	4,1	3,8	3,6	3,3	3,0	2,6	2,1
28	7,6	5,5	4,6	4,1	3,8	3,5	3,2	2,9	2,5	2,1
30	7,6	5,4	4,5	4,0	3,7	3,5	3,2	2,8	2,5	2,0
40	7,3	5,2	4,3	3,8	3,5	3,3	3,0	2,7	2,3	1,8
60	7,1	5,0	4,1	3,7	3,3	3,1	2,8	2,5	2,1	1,6
120	6,9	4,8	4,0	3,5	3,2	3,0	2,7	2,3	2,0	1,4
∞	6,6	4,6	3,8	3,3	3,0	2,8	2,5	2,2	1,8	1,0

Scanned with OKEN Scanner

*ANEXA 6**

DISTRIBUȚIA χ^2

Numărul gradelor de libertate	Probabilitatea				
	0,990	0,950	0,100	0,050	0,010
1			2,71	3,84	6,63
2	0,02	0,10	4,61	5,99	9,21
3	0,11	0,35	6,25	7,81	11,34
4	0,30	0,71	7,78	9,49	13,28
5	0,55	1,15	9,24	11,07	15,09
6	0,87	1,64	10,64	12,59	16,81
7	1,24	2,17	12,02	14,07	18,48
8	1,65	2,73	13,36	15,51	20,09
9	2,09	3,33	14,68	16,92	21,67
10	2,56	3,94	15,99	18,31	23,21
11	3,05	4,57	17,28	19,68	24,72
12	3,57	5,23	18,55	21,03	26,22
13	4,11	5,89	19,81	22,36	27,69
14	4,66	6,57	21,06	23,68	29,14
15	5,23	7,26	22,31	25,00	30,58
16	5,81	7,96	23,54	26,30	32,00
17	6,41	8,67	24,77	27,59	33,41
18	7,01	9,39	25,99	28,87	34,81
19	7,63	10,12	27,20	30,14	36,19
20	8,26	10,85	28,41	31,41	37,57
21	8,90	11,59	29,62	32,67	38,93
22	9,54	12,34	30,81	33,92	40,29
23	10,20	13,09	32,01	35,17	41,64
24	10,86	13,85	33,20	36,42	42,98
25	11,52	14,61	34,38	37,65	44,31
26	12,20	15,38	35,56	38,89	45,64
27	12,88	16,15	36,74	40,11	46,96
28	13,56	16,93	37,92	41,34	48,28
29	14,26	17,71	39,09	42,56	49,59
30	14,95	18,49	40,26	43,77	50,89
40	22,16	26,51	51,80	55,76	63,69
50	29,71	34,76	63,17	67,50	76,15
60	37,48	43,19	74,40	79,08	88,38
70	45,44	51,74	85,53	90,53	100,42
80	53,54	60,39	96,58	101,88	112,33
90	61,75	69,13	107,56	113,14	124,12
100	70,06	77,93	118,50	124,34	135,81

* După S N E D E C O R (1957)

Scanned with OKEN Scanner

VALORILE q ÎN %

t	Gradele de libertate ale erorii f																t
	1	2	3	4	5	6	7	8	9	10	12	15	20	30	50	∞	
1,8	32,3	21,3	17,0	14,7	13,2	12,2	11,5	10,9	10,5	10,2	9,5	9,1	8,7	8,1	7,6	7,2	1,8
1,9	30,8	19,8	15,4	13,0	11,6	10,7	9,9	9,3	9,0	8,7	8,0	7,6	7,2	6,6	6,1	5,7	1,9
2,0	29,6	18,4	13,9	11,6	10,2	9,3	8,5	8,0	7,7	7,3	6,8	6,4	6,0	5,4	4,9	4,5	2,0
2,1	28,3	17,1	12,7	10,4	9,0	8,0	7,4	6,9	6,6	6,3	5,8	5,3	4,8	4,4	4,0	3,6	2,1
2,2	27,2	15,9	11,5	9,3	8,0	7,0	6,4	6,0	5,6	5,3	4,8	4,4	4,0	3,6	3,2	2,8	2,2
2,3	26,1	14,8	10,5	8,3	7,0	6,1	5,5	5,1	4,7	4,4	4,1	3,6	3,2	2,9	2,5	2,1	2,3
2,4	25,1	13,8	9,6	7,4	6,2	5,3	4,7	4,3	4,0	3,7	3,4	3,0	2,6	2,3	2,0	1,6	2,4
2,5	24,2	13,0	8,8	6,7	5,4	4,6	4,1	3,7	3,4	3,1	2,8	2,5	2,1	1,8	1,6	1,2	2,5
2,6	23,4	12,1	8,0	6,0	4,8	4,1	3,5	3,2	2,9	2,7	2,4	2,0	1,7	1,4	1,23	0,93	2,6
2,7	22,7	11,4	7,4	5,4	4,3	3,6	3,1	2,7	2,5	2,3	2,0	1,7	1,4	1,1	0,98	0,70	2,7
2,8	21,9	10,7	6,8	4,9	3,8	3,1	2,6	2,3	2,1	1,9	1,6	1,4	1,1	0,90	0,74	0,52	2,8
2,9	21,2	10,1	6,3	4,4	3,4	2,7	2,3	2,0	1,8	1,6	1,3	1,1	0,89	0,70	0,57	0,38	2,9
3,0	20,5	9,7	5,8	4,0	3,0	2,4	2,0	1,7	1,5	1,33	1,1	0,91	0,71	0,55	0,43	0,27	3,0
3,1	19,9	9,1	5,4	3,6	2,7	2,12	1,7	1,45	1,25	1,12	0,93	0,75	0,57	0,43	0,33	0,20	3,1
3,2	19,3	8,6	5,0	3,3	2,4	1,86	1,5	1,25	1,08	0,97	0,78	0,60	0,45	0,33	0,25	0,14	3,2
3,3	18,8	8,1	4,6	3,1	2,2	1,65	1,3	1,08	0,93	0,82	0,65	0,50	0,36	0,26	0,18	0,10	3,3
3,4	18,3	7,7	4,3	2,75	2,0	1,45	1,13	0,95	0,79	0,70	0,53	0,40	0,28	0,20	0,14	<0,10	3,4
3,5	17,7	7,3	4,0	2,50	1,8	1,28	1,00	0,80	0,69	0,58	0,44	0,32	0,22	0,15	0,10		3,5
3,6	17,3	6,9	3,7	2,30	1,6	1,13	0,87	0,70	0,59	0,50	0,37	0,26	0,18	0,12	<0,10		3,6
3,7	16,8	6,6	3,4	2,10	1,4	1,01	0,77	0,61	0,50	0,43	0,30	0,21	0,15	<0,10			3,7
3,8	16,4	6,3	3,2	1,92	1,27	0,90	0,68	0,53	0,47	0,36	0,26	0,17	0,12				3,8
3,9	16,0	6,0	3,0	1,75	1,15	0,79	0,60	0,46	0,37	0,30	0,21	0,14	<0,10				3,9
	Diferențe nesemnificative		Diferențe semnificative			Diferențe distinct semnificative							Diferențe foarte semnificative				

Scanned with OKEN Scanner

4,0	15,6	5,7	2,8	1,62	1,05	0,71	0,52	0,39	0,31	0,26	0,18	0,12	4,0
4,2	14,8	5,2	2,45	1,35	0,84	0,57	0,40	0,29	0,23	0,18	0,13	<0,10	4,2
4,4	14,2	4,8	2,15	1,15	0,70	0,45	0,31	0,22	0,17	0,14	<0,10		
4,6	13,6	4,4	1,91	1,00	0,59	0,36	0,24	0,17	0,13	0,11			
4,8	13,0	4,1	1,70	0,86	0,49	0,29	0,19	1,13	<0,10	<0,10			
5,0	12,6	3,8	1,52	0,75	0,42	0,24	0,15	0,11					
5,2	12,1	3,5	1,37	0,66	0,34	0,20	0,12	<0,10					
5,4	11,7	3,25	1,23	0,58	0,29	0,17	0,10						
5,6	11,3	3,05	1,11	0,50	0,25	0,14	<0,10						
5,8	10,9	2,85	1,01	0,44	0,21	0,12							
6,0	10,5	2,65	0,92	0,39	0,18	0,10							
6,2	10,2	2,50	0,84	0,35	0,16	<0,10							
6,4	9,9	2,35	0,78	0,31	0,14								
6,6	9,7	2,22	0,72	0,27	0,12								
6,8	9,3	2,10	0,66	0,24	0,11								
7,0	9,0	1,99	0,61	0,22	<0,10								
8,0	7,9	1,52	0,41	0,13									
8,5	7,5	1,34	0,34	0,11									
9,0	7,0	1,19	0,29	<0,10									
9,5	6,7	1,08	0,25										
10,5	6,3	0,98	0,22										
13,0	4,90	0,58	0,10										
14,0	4,55	0,50	<0,10										

Cînd t experimental este mai mic decît 1,8 diferențele nu sînt semnificative.

* După SAULESCU (1959)

Scanned with OKEN Scanner

ANEXA 8

DIFERENȚE SEMNIFICATIVE TABELARE (Duncan *1955*)

Numărul gradelor de libertate	*p* — numărul de medii cuprinse în serie								
	2	3	4	5	6	7	8	9	10
	nivel de semnificație 0,05								
1	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0
2	6,09	6,09	6,09	6,19	6,09	6,09	6,09	6,09	6,09
3	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50
4	3,93	4,01	4,02	4,02	4,02	4,02	4,02	4,02	4,02
5	3,64	3,74	3,79	3,83	3,83	3,83	3,83	3,83	3,83
10	3,15	3,30	3,37	3,43	3,46	3,47	3,47	3,47	3,47
15	3,01	3,16	3,25	3,31	3,36	3,38	3,40	3,42	3,43
20	2,95	3,10	3,18	3,25	3,30	3,34	3,36	3,38	3,40
30	2,89	3,04	3,12	3,20	3,25	3,29	3,32	3,35	3,37
60	3,83	2,98	3,08	3,14	3,20	3,24	3,28	3,31	3,33
∞	2,77	2,92	3,02	3,09	3,15	3,19	3,23	3,26	3,29
	nivel de semnificație 0,01								
1	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0
2	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0
3	8,26	8,5	8,6	8,7	8,8	8,9	8,9	9,0	9,0
4	6,51	6,8	6,9	7,0	7,1	7,1	7,2	7,2	7,3
5	5,70	5,96	6,11	6,18	6,26	6,33	6,40	6,44	6,50
10	4,48	4,73	4,88	4,96	5,06	5,13	5,20	5,24	5,28
15	4,17	4,37	4,50	4,58	4,64	4,72	4,77	4,81	4,84
20	4,02	4,22	4,33	4,40	4,47	4,53	4,58	4,61	4,65
30	3,89	4,06	4,16	4,22	4,32	4,36	4,41	4,45	4,48
60	3,76	3,92	4,03	4,12	4,17	4,23	4,27	4,31	4,34
∞	3,64	3,80	3,90	3,98	4,04	4,09	4,14	4,17	4,20

Scanned with OKEN Scanner

ANEXA 9

TABEL CU NUMERE ÎNTÎMPLĂTOARE

(Mitropolski, *1961*)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1534	7106	2836	7873	5574	7545	7590	5574	1202	7712
6128	8993	4102	2551	0330	2358	6427	7067	9325	2454
6047	8566	8644	9343	9297	6751	3500	8754	2913	1258
0806	5201	5705	7355	1448	9562	7514	9205	0402	2427
9915	8274	4525	5695	5752	9630	7172	6988	0227	4264
2882	7158	4341	3463	1178	5786	1173	0670	0820	5067
9213	1223	4388	9760	6691	6861	8214	8813	0611	3131
8410	9836	3899	3683	1253	1683	6988	9978	8026	6751
9974	2362	2103	4326	3825	9079	6187	2721	1489	4216
3402	8162	8226	0782	3364	7871	4500	5598	9421	3816
8188	6596	1492	2139	8823	6878	0613	7161	0241	3834
3825	7020	1124	7483	9155	4919	3209	5959	2364	2555
9801	8788	6338	5899	3309	0807	0968	0539	4205	8257
5603	1251	6352	6467	0231	3556	2569	9446	4174	9219
0714	3757	0378	8266	8864	1374	6687	1221	0678	3714
4617	5652	7627	0372	8151	3668	1994	4402	2124	0016
6789	6279	7306	1856	7028	9043	7161	7526	6913	6393
6705	4978	8621	1790	4433	6298	0854	9127	3445	1111
3840	1086	0774	9241	9297	4233	1739	7734	0119	2436
7662	3939	2965	3273	0551	1645	8477	1877	5327	8629
7639	2868	4391	2950	7122	7325	9727	0080	7467	7947
3237	7203	4246	7329	7936	0065	4146	0866	4916	8648
3917	6271	1721	5469	1914	8653	0387	2756	6073	8984
9138	9395	6005	6423	7977	1873	7103	4267	9316	7206
8358	5896	6286	9242	5040	8509	2941	3913	3028	1563
1030	5094	1745	2975	2018	7340	6547	0207	5587	0300
6606	6305	1564	6668	7822	7142	6564	1659	5369	1659
4533	8841	4922	9365	1361	6692	1633	6774	0747	3881
4258	2012	0992	0106	1542	4760	0392	4057	0092	5203
5224	5128	8949	7928	7267	0116	1476	2009	1772	3860
6872	7492	7962	1867	7437	1526	3516	9129	4153	8084
8638	8407	7198	0956	0950	7753	5144	3914	5596	6104
9958	7172	5822	4224	6701	7559	4985	4856	4461	6147
0265	3086	2996	0699	3584	9702	1665	0446	9107	6437
8987	5441	7878	9404	0487	2939	3805	9172	7887	5197
5552	3529	9627	9362	6298	6021	0024	9520	9154	0643
9383	6640	7394	9592	9903	7699	8939	9972	1257	0994
9903	4059	0332	9109	0182	6721	9163	9008	2542	4461
6530	5070	7589	6928	6014	1832	9307	5107	1354	9257
8679	8953	8310	2060	6277	1773	7979	6741	6033	3588
5765	4987	1639	3512	9843	5286	3786	2384	4919	5611
7198	2447	6716	0291	5585	1106	5330	0504	6346	3679
2385	0605	2678	1399	2371	7968	1212	9569	8650	5841
0732	8732	8660	5836	9065	4603	0029	8042	0159	0345
1642	6094	3795	2600	4532	9740	0376	4384	9203	5387
4514	1956	7212	0687	7632	2106	0846	7055	4106	9157
8744	5580	8038	9087	7222	0424	0028	4511	3191	9846
3729	6225	5397	6790	2157	3414	6509	5204	4779	5641
8858	3147	8410	2783	1290	9796	8873	7585	7185	4726
3522	5601	6197	6051	3470	8283	5702	0103	8726	5282

Rîndurile sau coloanele vor fi folosite pe sărite; la fiecare repetiție vor fi folosite întotdeauna alte rînduri sau coloane.

Cinci variante vor fi așezate randomizat dacă se vor amplasa în următoarea ordine: 1, 5, 3, 4, 2 (după primul rînd) sau: 4, 5, 3, 2, 1 (după rîndul al treilea) etc.

Scanned with OKEN Scanner

BIBLIOGRAFIE

1. ASSMANN, E. — 1961 — Waldertragskunde, BLV Verlagsgesellschaft, München, Bonn, Wien.
2. BACKMAN, G. — 1943 — Wachstum und Organische Zeit. A. Barth, Leipzig.
3. BAILEY, N. — 1959 — Statistical methods in Biology (trad. în limba rusă), Moscova 1962.
4. BIJI, M., STOICHIȚA, I. — 1957 — Metoda selectivă în cercetarea selectivă. Editura Științifică, București.
5. BROWNLEE, K.A. — 1947 — Industrial experimentation (trad. în limba rusă: Statisticeskie metodî v proizvodstve, Moscova, 1949).
6. CEAPOIU, N. — 1959 — Elemente de tehnică experimentală agricolă, București.
7. COCHRAN, W.G. — 1953 — Sampling Techniques, New York.
8. COCHRAN, W.G., COX, G. — 1957 — Experimental designs, New York.
9. COSTEA, C. — 1962 — Codrul grădinărit, Editura Agro-Silvică, București.
10. COX, D.R. — 1961 — Planning of Experiments. New York.
11. CRISTESCU, V. ș.a. — 1957 — Stabilirea indicilor calitativi pentru 77 specii forestiere Analele I.C.E.S, vol. XVIII, Editura Agro-Silvică de Stat, București.
12. DIACICOV, A.N. — 1951 — Matematiceskaia statistica v primenenii k lesnomu delu. Arhanghelsk.
13. DISSESCU, R. — 1965 — Exploatarea mecano-grafică a datelor amenajistice, Revista Pădurilor, nr. 8.
14. DLIN, A.M. — 1958 — Matematiceskaia statistica v tehniche. Sovetskaia nauka, Moscova.
15. DORIN. T. — 1955 — Elemente de calcul statistic pentru silvicultori. Editura Agro-Silvică de Stat, București.
16. ERTELD, W. — 1958 — Richtlinien für die Anlage Bearbeitung von langfristigen waldbeulich — ertragskundlichen Versuchsflächen... Eberwalde.
17. FISHER, R.A. — 1954 — Statistical Methods for research workers. Edinburgh-London, Oliver and Boyd.
18. FISHER, R.A. — 1959 — Statistical Methods and Scientific Inference Edinburgh — London.
19. FRANZ, F. — 1965 — Auswertung von Dungungsversuchen mit unterschiedlicher Bestockungsdichte bei Versuchbeginn. Fortwiss. Cbl. 84, nr. 3—4.
20. GIURGIU, V. — 1955 — Studiul indicilor de formă la molidul din R.P.R. prin metoda statisticii variabile. Rev. Pădurilor, nr. 6.
21. GIURGIU, V. — 1956 — Spriamlenie crivoi vîsot. Naucino-teh. informația, nr. 17, M.L.T.I., Moscova.
22. GIURGIU, V. — 1957 — Taxația tecușevo prirosta nesajdenii. Dizertație. Manuscris, MLTI, Moscova.
23. GIURGIU, V. — 1964 — Principii și procedee pentru determinarea mărimii, creșterii și calității fondului de producție. Manuscris I.N.C.E.F.
24. GIURGIU, V. — 1965 — Algoritmi pentru calcule dendrometrice, C.D.F, București.
25. GIURGIU, R. — 1962 — Studiul taxatoric al tipurilor de pădure, Revista Pădurilor, nr. 4.
26. HALD, A. — 1952 — Statistical Theory with Engineering Applications, New York (Trad. în limba rusă: Matematiceskaia statistica s tehniceskimi prilojeniami, Moscova 1956).

Scanned with OKEN Scanner

27. HASSEL, A. A. — 1954 — Chapter 19 in Statistics in Mathematics and Biology, Iowa State College Press
28. HUSCH, B. — 1963 — Forest Mensuration and Statistics. The Ronald Press Company. New York.
29. I.U.F.R.O. — 1959 — La normalisation des symboles dans les mesures forestières.
30. I.U.F.R.O. — 1965 — Revised list of papers submitted to the conference Advisory grup of Forest statisticans of the I.U.F.R.O. Stockholm.
31. JEFFERS, J.—1953—Statistical Methods in Forest Research. I.U.F.R.O. 11 Congr. Rome.
32. JEFFERS, J. — 1960 — Experimental Design and analysis in Forest Research, Stockholm.
33. JEFFERS, J. — 1963 — Report of the international advisory grup of forest statisticians. Forest Comission: Forest Record 56
34. KEMPTHORNE, O. — 1963 — An Introduction to Genetic Statistics New York.
35. LEONTIEV, N. L. — 1961 — Tehnica statisticeskih vîcislenii. Goslesbumizdat — 1961.
36. LINDER, A. — 1960 — Statistiche Methoden für Naturwissenschaler, Mediziner und Ingenieure. Basel.
37. LOETSCH-HALLER — 1964 — Forest Inventory. BLV Verlagsgesellschaft München, Basel, Wien.
38. MANEA, M (red.) — 1964 — Calcul economic. Editura Academiei R.P.R., București.
39. MIHOC, Gh., URSEANU, V. — 1962 — Matematici aplicate în statistică, Editura Academiei R.P.R., București.
40. MITROPOLSKI, A. K. — 1961 — Tehnica statisticeskih vîcislenii. Fizmatghiz. Moscova
41. MUDRA, A. — 1958 — Statistiche Methoden für landwirtschaftiche Versuche, Berlin.
42. MUNTEANU, N. — 1964 — Metode neparametrice folosite în statistică. Revista de statistică, nr. 11.
43. NALIMOV, V. V. — 1960 — Primenenie matematiceskoi statistiki pri analize veșestva. Fizmatghiz. Moscova.
44. ONICESCU, O., MIHOC, Gh. — 1958 — Lecții de statistică matematică. Editura Tehnică, București.
45. PARDÉ, J. — 1961 — Dendrometrie, Nancy.
46. POPESCU-ZELETIN, I. — 1947 — Procedeul de inventariere cu benzi de probă, Rev. Pădurilor, nr. 1 — 3, 4 — 6.
47. PRODAN, M. — 1961 — Forstliche Biometrie. BVL Verlagsgesellschaft. München, Bonn, Wien.
48. QUENQUILLE, M. H. — 1953 — The Design and Analisys of experiment. London.
49. RANCU, N., TÖVISSI, L. — 1963 — Statistica matematică cu aplicații în producție. Editura Academiei R.P.R., București.
50. RANCU, N., TÖVISSI. L. — 1964 — Analiza statistico-matematică a calității producției industriale. Editura Științifică, București.
51. THOMASIUS, H. — 1964 — Beitrag zur Durchführung forstlicher Düngungsversuche. Wiss. Z. Tech. Univ. Dresden, 13, nr. 5.
52. SĂHLEANU, V. — 1957 — Metode matematice în cercetarea medico-biologică. Editura Medicală, București.
53. SĂULESCU, N. — 1959 — Cîmpul de experiență. Editura Agro-silvică, București.
54. SCATTERGOOD, L. W. — 1954 — Chapter 20 in Statistics in Mathematics and Biology. Iowa State College Press.
55. SCHUMACHER, F. X., CHAPMANN. — 1954 — Sampling methods in forestry and range management. Durham, North Carolina.
56. SNEDECOR, G. — 1957 — Statistical Methods applied to experiments in agriculture and biology. The Iowa State College Press. Ames Iowa (tr. în limba rusă, Moscova 1961).
57. STEINBACH, M. — 1961 — Prelucrarea statistică în medicină și biologie. Editura Academiei R.P.R., București
58. TOMASSONE, R. — 1964 — L'analyse de covariance. Ann. des sciences forestieres, Nancy.
59. TREBICI, V., RANCU, N. — 1964 — Aplicații ale analizei secvențiale la verificarea ipotezelor statistice. Calcul economic. Editura Academiei R.P.R., București.

Scanned with OKEN Scanner

Scanned with OKEN Scanner

TABLA DE MATERII

I. Introducere 3
1.1 Generalităţi 3
1.2 Observarea şi gruparea statistică 6
1.3 Inregistrarea şi prelucrarea iniţială a datelor 7
1.4 Reprezentări grafice ale distribuţiilor statistice 9

II. Mărimi medii 14
2.1 Generalităţi 14
2.2 Momente 14
2.3 Media aritmetică, media pătratică şi media geometrică 21
2.4 Mediana, qartile, decile, modul 22
2.5 Indicatorii variaţiei 25

III. Distribuţii statistice frecvent întîlnite în cercetările din domeniul silviculturii . . . 33
3.1 Distribuţia binomială 33
3.2 Distribuţia evenimentelor rare sau distribuţia Poisson 36
3.3 Distribuţia normală 39
3.4 Asimetria şi excesul 46
3.5 Ajustarea distribuţiilor asimetrice 52
3.6 Erorile principalilor parametri ai distribuţiilor statistice unidimensionale 54

IV. Exemple din silvicultura privind verificarea ipotezelor statistice 56
4.1 Consideraţii generale 56
4.2 Compararea a două medii 59
4.3 Compararea dispersiilor 67
4.4 Verificarea concordanţei dintre distribuţia empirică şi distribuţia teoretică . . 69

V. Analiza dispersională cu aplicaţii în silvicultură 74
5.1 Generalităţi 74
5.2 Analiza dispersională în cazul unei grupări simple 75
5.3 Analiza dispersională în cazul a doi factori 84
5.4 Analiza dispersională pentru experienţe polifactoriale 92
5.5 Aşezarea experienţelor 97
5.6 Analiza covarianţei 104

VI. Despre corelaţie 108
Generalităţi 108
6.2 Coeficientul de corelaţie 111
6.3 Stabilirea semnificaţiei unei corelaţii 118

Scanned with OKEN Scanner

6.4 Corelația multiplă 124
6.5 Corelația parțială 126
6.6 Corelația rangurilor 126
6.7 Tabele de contingență 127

VII. Exemple de ecuații de regresie din biometria forestieră 132

7.1 Generalități 132
7.2 Câteva considerații generale privind alegerea tipului ecuației de regresie . . . 133
7.3 Estimarea coeficienților din ecuațiile de regresie 137
7.4 Metoda celor mai mici pătrate 139
7.5 Raportul de corelație 152

VIII. Metoda selectivă și importanța ei pentru silvicultură 157

8.1 Introducere 157
8.2 Selecția întîmplătoare nerepetată 160
8.3 Selecția întîmplătoare repetată 166
8.4 Selecția stratificată nerepetată 166
8.5 Selecția mecanică 171
8.6 Sondaje în două sau mai multe faze 172

Anexa 1: Valorile $x\,n_j$ 179
Anexa 2: Valorile funcției de frecvență ale distribuției normale $f(u)$ și ale derivatelor $f^{III}(u)$ și $f^{IV}(u)$ 180
Anexa 3: Valorile probabilității $q = 1 - 2\varphi(u)$ corespunzătoare distribuției normale . . . 182
Anexa 4: Valorile distribuției Student 183
Anexa 5: Distribuția F 184
Anexa 6: Distribuția χ^2 187
Anexa 7: Valorile Q în % 188
Anexa 8: Diferențe semnificative tabelare 190
Anexa 9: Tabel cu numere întîmplătoare 191

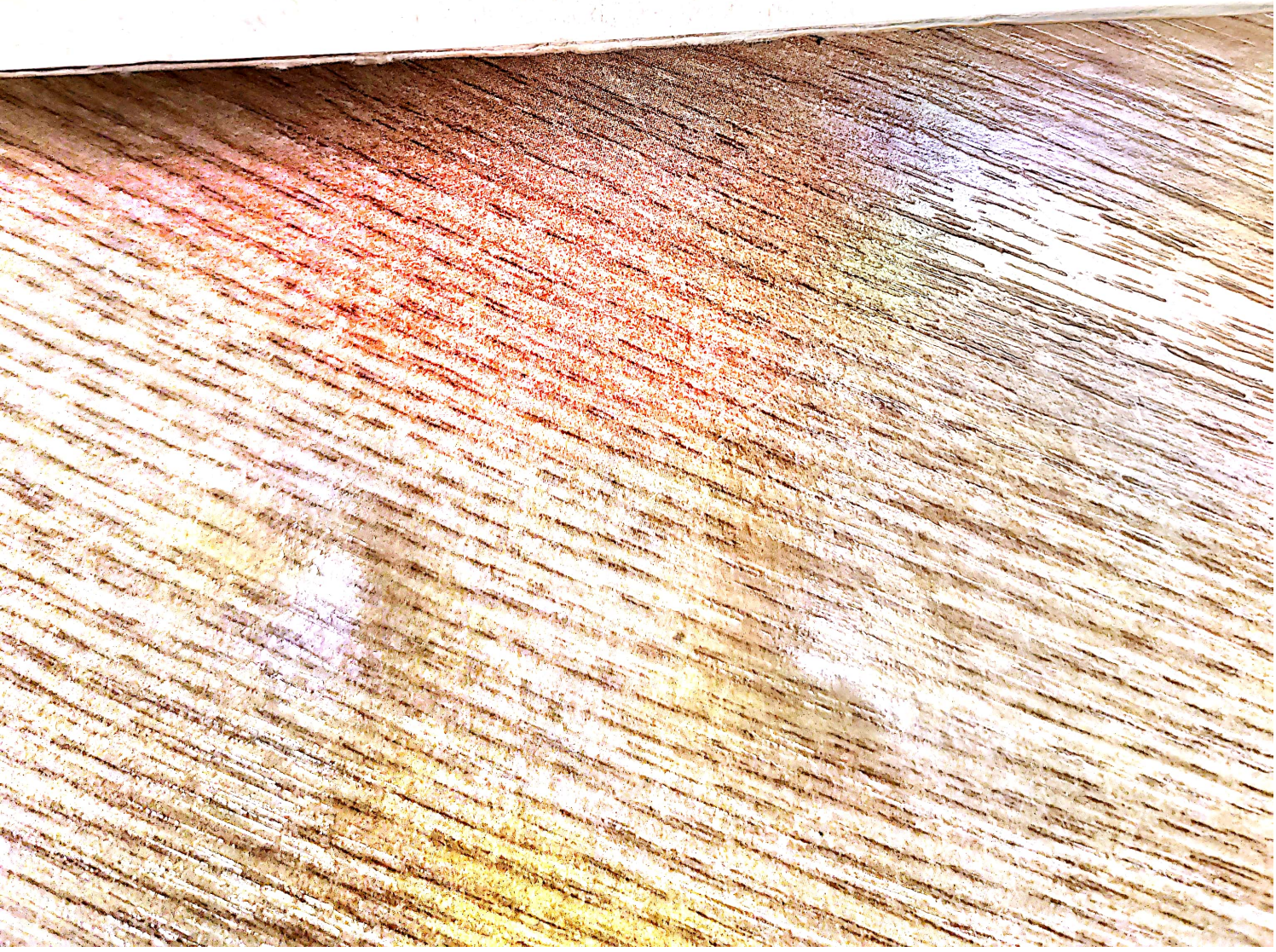
Bibliografie 193

Scanned with OKEN Scanner

Redactor responsabil: Ing. CRĂCEA AURORA RODICA
Tehnoredactor: TOMA ALEXANDRESCU

Dat la cules 05.04.1966. Bun de tipar 01.08.1966. Apărut 1966
Tiraj 800 + 100 ex. Hîrtie tipar înalt tip A 80 g/m², ft. 70×100/16
Coli editoriale 11,60. Coli de tipar 12,50 planșe 1. A 4707. C.Z.
pentru bibliotecile mari 519.2:634.9 C.Z. pentru bibliotecile mici 51:634

Întreprinderea Poligrafică „Tiparul", str. Fabrica de chibrituri nr. 9-11, București — Republica Socialistă România.
Comanda nr. 225



Scanned with OKEN Scanner